Removal Lemmas: Summer School 2025

1 The regularity and removal lemmas

The graph removal lemma is the following statement:

Theorem 1.1 (Graph removal lemma, Ruzsa-Szemerédi '78)**.** *Let H be a fixed graph. For every $\varepsilon > 0$ there is $\delta = \delta_H(\varepsilon) > 0$ such that if an n-vertex graph G has at most $\delta n^{v(H)}$ copies of H, then G can be made H-free by deleting at most εn^2 edges.*

Remarks:

- We say that G is *ε-far* from being H-free if one has to delete at least εn^2 edges to turn G into an H-free graph. The contrapositive is that if G is ε-far from H-free then G has at least $\delta n^{v(H)}$ copies of H.
- Being ε-far from H-free is equivalent to having a collection of $\Theta(\varepsilon)n^2$ edge-disjoint copies of H. Indeed, if G has such a collection of size εn^2, then G is ε-far (because we have to delete at least one edge from each H-copy in order to destroy all H-copies in G). In the other direction, take a maximal collection of edge-disjoint copies of H in G. Deleting all edges of these copies makes the graph H-free (because of the maximality of the collection). Thus, if the maximal such collection has size less than $\frac{\varepsilon}{e(H)}n^2$, then G is not ε-far.

The removal lemma is proved using Szemerédi's regularity lemma, which we now recall. Consider a bipartite graph with parts X, Y. The *density* is $d(X,Y) := \frac{e(X,Y)}{|X||Y|}$.

Definition 1.2 (Regular pair)**.** *A bipartite graph (X, Y) is* ε-regular *if for every $X' \subseteq X, Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|, |Y'| \geq \varepsilon|Y|$, it holds that $|d(X', Y') - d(X,Y)| \leq \varepsilon$.*

Regular pairs are "random-like". Indeed, the definition captures a key property of random graphs: uniform edge distribution. Another key random-like property of regular pairs is given by the counting lemma:

Lemma 1.3 (Counting lemma)**.** *For every $\varepsilon > 0$ there is $\gamma > 0$ such that if $V_1, \dots, V_r$ are disjoint vertex sets such that all pairs (V_i, V_j) are ε-regular, then the number of r-cliques $v_1, \dots, v_r$ (with $v_i \in V_i$) is*

$$\prod_{i=1}^{r} |V_i| \cdot \left(\prod_{1 \leq i < j \leq r} d(V_i, V_j) \pm \gamma \right). \tag{1}$$

Note that (1) (with the error γ omitted) is precisely the expected number of r-cliques if the edges between V_i and V_j were chosen randomly with probability $d(V_i, V_j)$, for every $1 \leq i < j \leq r$. In many applications, it suffices to have a lower bound for the number of r-cliques. To illustrate how the proof of the counting lemma works, let us prove such a statement in the case $r = 3$ (the proof for general r is similar, via induction). We will assume that all densities $d(V_i, V_j)$ are large enough in terms of ε.[1]

Lemma 1.4. *For every $d > 0$ there is $\varepsilon = d/2$ so that if V_1, V_2, V_3 are such that $d(V_i, V_j) \geq d$ and (V_i, V_j) is ε-regular for every $1 \leq i < j \leq 3$, then there are at least $(d^3 - 4\varepsilon)|V_1||V_2||V_3|$ triangles.*

Proof. First we need the following simple property of regular pairs. The proof is left to the reader.

Claim 1.5. *Let (X, Y) be an ε-regular pair with density $d = d(X, Y)$. Then at most $\varepsilon|X|$ of the vertices $x \in X$ satisfy $\frac{d_Y(x)}{|Y|} < d - \varepsilon$, and at most $\varepsilon|X|$ of the vertices $x \in X$ satisfy $\frac{d_Y(x)}{|Y|} > d + \varepsilon$.*

Now we prove Lemma 1.4. For $i = 1, 2$, let B_i be the set of vertices $v \in V_3$ with $d_{V_i}(v) < (d-\varepsilon)|V_i|$. By Claim 1.5, we have $|B_i| \leq \varepsilon|V_3|$. So $|B_1 \cup B_2| \leq 2\varepsilon|V_3|$. For each $v \in V_3 \setminus (B_1 \cup B_2)$, consider $U_1 := N_{V_1}(v)$ and $U_2 := N_{V_2}(v)$. As $v \notin B_1 \cup B_2$, we have $|U_1| \geq (d - \varepsilon)|V_1| \geq \varepsilon|V_1|$ and similarly $|U_2| \geq \varepsilon|V_2|$. By the regularity of (V_1, V_2), we have $d(U_1, U_2) \geq d - \varepsilon$, and therefore $e(U_1, U_2) \geq (d - \varepsilon)|U_1||U_2| \geq (d - \varepsilon)^3|V_1||V_2|$. Each edge in $E(U_1, U_2)$ creates a triangle with v. Doing this for all (at least $(1 - \varepsilon)|V_3|$) choices of $v \in V_3 \setminus (B_1 \cup B_2)$, we get at least $(1 - \varepsilon)(d - \varepsilon)^3|V_1||V_2||V_3| \geq (d^3 - 4\varepsilon)|V_1||V_2||V_3|$ triangles, as required. ■

Another version of the counting lemma we will use is as follows.

Definition 1.6 (bi-induced copy)**.** *A* bi-induced copy *of a bipartite graph $H = (A, B)$ in a graph G is an injection $\varphi : V(H) \to V(G)$ such that for every $a \in A, b \in B$, $ab \in E(H)$ if and only if $\varphi(a)\varphi(b) \in E(G)$. If G is itself bipartite with parts X, Y, then we also require that $\varphi(A) \subseteq X$ and $\varphi(B) \subseteq Y$.*

Note that in the above definition we do not make requirements on the edges inside $\varphi(A)$ and $\varphi(B)$.

Lemma 1.7. *For every integer k and $d > 0$, there is $\varepsilon > 0$ such that the following holds. Consider a bipartite graph (X, Y) and suppose that $d \leq d(X, Y) \leq 1 - d$ and (X, Y) is ε-regular. Then (X, Y) contains a bi-induced copy of every bipartite graph (A, B) with $|A|, |B| \leq k$.*

One can deduce the above lemma from Lemma 1.3 as follows: Suppose that $A = \{a_1, \dots, a_k\}$, $B = \{b_1, \dots, b_k\}$. Split X into equal parts $X_1, \dots, X_k$ and Y into equal parts $Y_1, \dots, Y_k$. Define an auxiliary graph as follows: If $a_i b_j \in E$ then take the edges of G between X_i, Y_j, and if $a_i b_j \notin E$ then take the non-edges of G between X_i, Y_j. Now apply Lemma 1.3 to this auxiliary graph.

The Szemerédi regularity lemma states that any graph has a vertex partition into a bounded number of parts, such that most pairs of parts are regular.

Theorem 1.8 (Szemerédi's regularity lemma 1978)**.** *For every $\varepsilon > 0$ and $t_0 \geq 1$, there is $T = T(\varepsilon, t_0)$ such that the following holds. Every graph G on $n \geq T$ vertices has an equipartition[2] $V(G) = V_1 \cup \dots \cup V_t$ with $t_0 \leq t \leq T$ such that all but εt^2 of the pairs (V_i, V_j), $1 \leq i < j \leq t$, are ε-regular.*

An equipartition as in Theorem 1.8 is called *ε-regular*. Let us give a very rough sketch of the proof of the regularity lemma.

---

[1] Otherwise, i.e. if some $d(V_i, V_j)$ is smaller than γ, then it is easy to see that the statement of Lemma 1.3 holds trivially (because the number of r-cliques is at most $|V_1| \dots |V_r| d(V_i, V_j)$).

[2] An equipartition is a partition in which any two parts V_i, V_j satisfy $||V_i| - |V_j|| \leq 1$.

Proof sketch of Theorem 1.8. For a partition $\mathcal{P} = \{V_1, \dots, V_t\}$ of $V(G)$, we define the *mean square density* as

$$q(\mathcal{P}) = \sum_{1 \le i < j \le t} \frac{|V_i||V_j|}{n^2} d^2(V_i, V_j).$$

One shows that:

1. If $\mathcal{Q}$ is a refinement of $\mathcal{P}$ then $q(\mathcal{Q}) \ge q(\mathcal{P})$.
2. If $\mathcal{P}$ is not ε-regular, then there is a refinement $\mathcal{Q}$ of $\mathcal{P}$ with $q(\mathcal{Q}) \ge q(\mathcal{P}) + \varepsilon^5$ and $|\mathcal{Q}| \le 2^t \cdot t$, where $t = |\mathcal{P}|$.

A proof sketch for Item 2 is as follows: For each pair $1 \le i < j \le t$ such that (V_i, V_j) is not ε-regular, take $V_{ij} \subseteq V_i, V_{ji} \subseteq V_j$ such that $|V_{ij}| \ge \varepsilon|V_i|$, $|V_{ji}| \ge \varepsilon|V_j|$, and $|d(V_{ij}, V_{ji}) - d(V_i, V_j)| > \varepsilon$. Now, for each $1 \le i \le t$, take the common refinement (Venn diagram) of all sets $(V_{ij} : j)$. The resulting partition is Q. It is easy to see that $|\mathcal{Q}| \le 2^t \cdot t$, and one can show that $q(\mathcal{Q}) \ge q(\mathcal{P}) + \varepsilon^5$.

Using Items 1-2, one proves Theorem 1.8 as follows. Start with an arbitrary equipartition $\mathcal{P}_0$ into t_0 parts. If $\mathcal{P}_i$ is not ε-regular, use Item 2 to get a refinement $\mathcal{P}_{i+1}$ with $q(\mathcal{P}_{i+1}) \ge q(\mathcal{P}_i) + \varepsilon^5$. As $q(\mathcal{P}) \le 1$ for any $\mathcal{P}$, the process has to stop in at most $\frac{1}{\varepsilon^5}$ steps.

At each iteration, there is also an additional step of turning the partition $\mathcal{Q}$ given by Item 2 into an equipartition, by chopping up the parts of $\mathcal{Q}$ into equal-sized sets. One can show that if the set-size is small enough, this does not decrease $q(\mathcal{Q})$ by much. ■

What is the bound on the partition-size T that we get in Theorem 1.8? The proof is via a procedure that runs for poly$(1/\varepsilon)$ steps, and at each step we replace a partition of size t with a partition of size roughly 2^t. Hence, the number of parts is at most tower(poly$(1/\varepsilon), t_0)$, where

$$\text{tower}(k, x) = 2^{2^{2^{\cdot^{\cdot^{\cdot^{2^x}}}}}} \Big\} k \text{ times}$$

I.e., the bound is of tower type. Gowers proved that this is inevitable.

Theorem 1.9 (Gowers 1997)**.** *There are graphs which require* tower(ε^{-c}) *parts in any ε-regular partition, where $c > 0$ is a constant.*

Let us now prove the removal lemma (Theorem 1.1). For simplicity, we consider the case $H = K_3$.

Proof of the triangle removal lemma. Let G be a graph which is γ-far from K_3-free. Apply the regularity lemma with parameters $\varepsilon = \gamma/10$ and $t_0 = 10/\gamma$ to obtain an ε-regular partition $V_1, \dots, V_t$ with $t_0 \le t \le T$. We now clean the graph. I.e., we delete the following edges:

1. All edges inside V_i for every $1 \le i \le t$.
2. All edges between pairs (V_i, V_j) with $d(V_i, V_j) \le 2\varepsilon$.
3. All edges between pairs (V_i, V_j) which are not ε-regular.

The number of edges of type 1 is at most $t \cdot \binom{n/t}{2} \le \frac{n^2}{t} \le \frac{\gamma}{10}n^2$. The number of edges of type 2 is at most $2\varepsilon \cdot \sum_{1 \le i < j \le t} |V_i||V_j| \le 2\varepsilon\binom{n}{2} \le \frac{\gamma}{5}n^2$. The number of edges of type 3 is at most $\varepsilon t^2 \cdot \left(\frac{n}{t}\right)^2 = \varepsilon n^2 = \frac{\gamma}{10}n^2$. So the total number of deleted edges is less than γn^2. As G is γ-far from K_3-free, the remaining graph (after the deletion of these edges) still has a triangle. This triangle cannot contain two vertices from the same part V_i (because of Item 1). So suppose that this triangle has one vertex in each of the sets V_i, V_j, V_k. Then by Items 2-3, all pairs $(V_i, V_j), (V_i, V_k), (V_j, V_k)$ are ε-regular and have density at least 2ε. By Lemma 1.4, there are at least $\text{poly}(\varepsilon)|V_i||V_j||V_k|$ triangles. Now, $|V_i||V_j||V_k| = (n/t)^3 \geq n^3/T^3$, so we can set $\delta := \frac{\text{poly}(\varepsilon)}{T^3}$. ■

Note that because of the tower-type parameter dependence in the regularity lemma, the above proof gives a tower-type dependence for the removal lemma as well. Namely, it shows that in Theorem 1.1, we can take $1/\delta = \text{tower}(\text{poly}(1/\varepsilon))$. This was improved to $\text{tower}(O(\log \frac{1}{\varepsilon}))$ by Fox. It is a major open problem to improve this further.

2 When is the removal lemma polynomial?

For which graphs H does it hold that the parameters in the H-removal lemma (Theorem 1.1) satisfy $\delta_H(\varepsilon) = \text{poly}(\varepsilon)$? A classical result in extremal graph theory, namely the Kővári-Sós-Turán theorem, shows that this is the case if H is bipartite.

Theorem 2.1 (Kővári-Sós-Turán theorem, supersaturation form)**.** *An n-vertex graph with εn^2 edges contains at least poly(ε)n^{s+t} copies of $K_{s,t}$.*

Returning to the H-removal lemma for a bipartite H, observe that if G is ε-far from H-free then G (trivially) contains at least εn^2 edges, hence G contains $\text{poly}(\varepsilon)n^{v(H)}$ copies of H by Theorem 2.1. Thus, if H is bipartite then the H-removal lemma is polynomial. Alon proved that the converse also holds, i.e., that bipartite graphs are the only ones which admit a polynomial removal lemma.

Theorem 2.2 (Alon 2002)**.** *For a graph H, $\delta_H(\varepsilon) = \text{poly}(\varepsilon)$ if and only if H is bipartite.*

We will first prove Theorem 2.2 in the case that H is an odd cycle. For this, we need a number-theoretic construction.

Theorem 2.3. *Let $k \geq 3$. There is a set $S \subseteq [n]$ with $|S| \geq n^{1-o(1)}$, such that for every $x_1, \dots, x_k \in S$, if $x_1 + \dots + x_{k-1} = (k-1)x_k$, then $x_1 = \dots = x_k$.*

The case $k = 3$ is Behrend's construction of a large set with no 3-term arithmetic progressions. The general case is a straightforward generalization.

Proof of Theorem 2.3. Write $n = d^t$ for d, t to be chosen later. Represent the numbers $1, \dots, n$ in base d. I.e., for $x \in [n]$, write

$$x = \sum_{i=0}^{t-1} a_i d^i,$$

where $0 \leq a_i \leq d-1$. Write $v(x) := (a_0, \dots, a_{t-1})$. Let U be the set of all x for which $a_0, \dots, a_{t-1} \leq \frac{d-1}{k-1}$. This property guarantees that for $x_1, \dots, x_{k-1} \in U$, we have

$$v(x_1 + \dots + x_{k-1}) = v(x_1) + \dots + v(x_{k-1}).$$

I.e., there is no carry when summing $x_1, \dots, x_{k-1}$. Similarly, $v((k-1) \cdot x_k) = (k-1) \cdot v(x_k)$ for every $x_k \in U$.

Now fix $r \geq 1$, to be chosen later, and take S to be the set of all $x \in U$ with $\|v(x)\| = r$, where $\|\cdot\|$ is the Euclidean norm. Suppose that $x_1, \dots, x_k \in S$ satisfy $x_1 + \dots + x_{k-1} = (k-1)x_k$. Putting

$v_i = v(x_i)$, we get $v_1 + \cdots + v_{k-1} = (k-1)v_k$. Now we take norms. The norm of the RHS is $(k-1)r$. For the LHS, by Cauchy-Schwarz we have $\|v_1 + \cdots + v_{k-1}\| \leq \sqrt{\sum_{i=1}^{k-1} \|v_i\|^2} \cdot \sqrt{k-1} = (k-1)r$, with equality if and only if $v_1 = \cdots = v_{k-1}$. So we must have $v_1 = \cdots = v_k$ and hence $x_1 = \cdots = x_k$.[3]

Now we estimate the size of S. For every $x \in [n]$, we have $\|x\|^2 \leq td^2$, so the number of choices for r is at most td^2. By pigeonhole, there exists r such that

$$|S| \geq \frac{|U|}{td^2} \geq \frac{(d/k)^t}{td^2} = \frac{n}{k^t td^2}.$$

Choose t, d such that $k^t = d$. As $d^t = n$, this gives $t = \sqrt{\frac{\log(n)}{\log(k)}}$, $d = e^{\sqrt{\log(k)\log(n)}}$. So

$$|S| \geq \frac{n}{e^{O_k(\sqrt{\log n})}} = n^{1-o(1)}.$$

■

Now we prove Theorem 2.2 for odd cycles.

Theorem 2.4. *For every odd $k \geq 3$, there exists an n-vertex graph G with εn^2 edge-disjoint copies of C_k, but only $\varepsilon^{\omega(1)} n^k$ copies of C_k in total.*

Proof. Let $\varepsilon > 0$. Let $S \subseteq [n]$ be the set given by Theorem 2.3. Choose n such that $|S| = \varepsilon n$. As $|S| = n^{1-o(1)}$, this means that $n = (1/\varepsilon)^{\omega(1)}$. Define a graph with k parts $V_1, \dots, V_k$, each of size kn and identified with $[kn]$.[4] For each $y \in [n]$ and $x \in S$, add a copy of C_k on the vertices $v_1 = y, v_2 = y + x, v_3 = y + 2x, \dots, v_k = y + (k-1)x$ (so $v_i = y + (i-1)x$) such that $v_i \in V_i$.[5] Denote this copy by $C_{x,y}$. We claim that the copies $C_{x,y}$ are edge-disjoint. Indeed, even stronger, any two such copies share at most one vertex, because if $C_{x,y}$ and $C_{x',y'}$ have the same vertex in V_i and V_j, then $y + (i-1)x = y' + (i-1)x'$ and $y + (j-1)x = y' + (j-1)x'$, and solving this system of equations gives $x = x', y = y'$. The number of copies $C_{x,y}$ is $n|S| \geq \varepsilon n^2$. Thus, the graph has a collection of εn^2 edge-disjoint copies of C_k.

Now we bound the total number of copies of C_k. Crucially, as k is odd, we can only have copies of C_k of the form $(v_1, \dots, v_k, v_1)$ with $v_i \in V_i$.[6] Now consider such a copy $v_1, \dots, v_k$. Then for each $1 \leq i \leq k-1$ there are y_i, x_i with $v_i, v_{i+1} \in C_{x_i,y_i}$, and there are y_k, x_k with $v_k, v_1 \in C_{x_k,y_k}$. Then

$$x_1 + \cdots + x_{k-1} = v_k - v_1 = (k-1)x_k.$$

By the property of the set S, we get $x_1 = \cdots = x_k =: x$ (from which we can also deduce that $y_1 = \cdots = y_k$). So $(v_1, \dots, v_k) \in C_{x,y_1}$. This shows that any copy of C_k in the graph is one of the "original" copies $C_{x,y}$ we put in. Their number is

$$n|S| \leq n^2 \leq \frac{|V(G)|^k}{n} \leq \varepsilon^{\omega(1)} |V(G)|^k.$$

■

[3] What we are using here is that S is a sphere, and a sphere has no point in the convex hull of other points (unless all points are equal.

[4] Thus, we are actually defining a graph on $k^2 n$ vertices, but we can of course adjust the parameters.

[5] Note that we choose each V_i to be $[kn]$ so that the numbers $v_i = y + (i-1)x$ "fit" in V_i.

[6] What we are using here is that C_k is not homomorphic to any of its proper subgraphs.

Remarks:

- We can take blowups of the graph defined in the proof of Theorem 2.4 to get constructions of any (large enough) size.
- The above proof gives a connection between the triangle removal lemma and the problem of estimating the largest possible size $r_3(n)$ of a subset of $[n]$ with no 3-term arithmetic progression. Indeed, in the proof, we use a lower bound on $r_3(n)$ (via Theorem 2.3) to show that the triangle removal lemma is not polynomial. In the other direction, one can use the triangle removal lemma to show that $r_3(n) = o(n)$, which is the statement of Roth's theorem. We note, however, that this gives a very poor quantitative bound of roughly $r_3(n) \leq n/\log_*(n)$. Much better bounds are known.

To prove Theorem 2.2 for a general non-bipartite H, we would like to use the same strategy as in Theorem 2.4. Namely, if $V(H) = \{1, \dots, h\}$, we construct an H-partite graph with parts $V_1, \dots, V_h$ and put a copy of H on $y, y + x, \dots, y + (h-1)x$ for $y \in [n], x \in S$. We will also use that H has an odd cycle. The issue is that we want to make sure that every copy of H is of the form $v_1, \dots, v_h$ with $v_i \in V_i$ (and v_i plays the role of $i \in [h]$). Note that the construction is homomorphic to H via the homomorphism $V_i \mapsto i$.[7] Thus, what we want is that H has no homomorphism to a proper subgraph of itself. This might not be true of H itself, but there is a maximal subgraph of H which has this property, and we will exploit this for our construction. Let us now define this subgraph.

Definition 2.5. *The* core *of H is the minimal subgraph K of H (in terms of the number of vertices) such that there is a homomorphism from H to K.*

We will show soon that the core is well defined, in the sense that K is unique up to isomorphism. Observe that K is not homomorphic to any of its proper subgraphs. Indeed, if there is a homomorphism $\psi : K \to J$ for J with $V(J) \subsetneq V(K)$, then by taking a homomorphism $\varphi : H \to K$, we get a homomorphism $\psi \circ \varphi$ from H to J, contradicting the minimality of K. Thus, every homomorphism from K to itself is injective and hence an isomorphism. Similarly, we can show that the core is unique up to isomorphism: If K_1, K_2 are both cores of H, then there are homomorphisms $\varphi_1 : K_2 \to K_1$ and $\varphi_2 : K_1 \to K_2$ (we obtain φ_i by taking a homomorphism from H to K_i and restricting it to K_{3-i}). Now, $\varphi_1 \circ \varphi_2$ is a homomorphism from K_1 to itself and hence an isomorphism, and similarly for $\varphi_2 \circ \varphi_1$. It follows that φ_1, φ_2 are bijective and hence isomorphisms.

Note that if H is bipartite (and has at least one edge), then the core of H is an edge. On the other hand, if H is not bipartite then neither is its core. Using cores, we can now prove Theorem 2.2. The idea is to do the construction for the core of H, and then blow it up by a constant factor to get a construction for H.

Proof of Theorem 2.2. Let K be the core of H. Then K is also not bipartite. Write $V(K) = \{1, \dots, k\}$, where $(1, \dots, \ell, 1)$ is an odd cycle. Take $S \subseteq [n]$ from Theorem 2.3 (with parameter ℓ), and define a graph G with sides $V_1, \dots, V_k$ by doing the following: For each $y \in [n]$ and $x \in S$, put a copy $K_{x,y}$ of K on $v_1, \dots, v_k$, where $v_i = y + (i-1)x \in V_i$ (in this copy, v_i plays the role of i). As in the proof of Theorem 2.4, the copies $K_{x,y}$ are edge-disjoint, and hence G has εn^2 edge-disjoint copies of K.

On the other hand, since K is a core, every copy of K in G is of the form $v_1, \dots, v_k$ with $v_i \in V_i$ playing the role of i. Hence, for each such copy $v_1, \dots, v_k$, the vertices $v_1, \dots, v_\ell$ makes an odd

[7] A *homomorphism* from a graph G to a graph H is a mapping $\varphi : V(G) \to V(H)$ such that $\varphi(x)\varphi(y) \in E(H)$ for every $xy \in E(G)$.

cycle. By the same argument as in the proof of Theorem 2.4, each such odd cycle is of the form $(y, y + x, \ldots, y + (\ell - 1)x)$ for some $y \in [n], x \in S$, and hence the number of such odd cycles is at most n^2. Thus, the total number of copies of K in G is at most $n^2 \cdot n^{k-\ell} \leq n^k/n \leq \varepsilon^{\omega(1)} n^k$.

To obtain a construction for H, take the above construction for K and blow it up by a factor of $h := |V(H)|$. Then each copy $K_{x,y}$ of K gives rise to a copy of H (because H is homomorphic to K, i.e., contained in a blowup of K). Hence, the resulting graph (which has $O(n)$ vertices) has εn^2 edge-disjoint copies of H. On the other hand, each copy of H must contain a copy of K (because K is a subgraph of H). Also, the blown-up graph has $O(\varepsilon^{\omega(1)} n^k)$ copies of K (the only way to get copies of K is from blowups of copies of K in G, as K is a core). Thus, the total number of copies of H is $O(\varepsilon^{\omega(1)} n^k) \cdot n^{h-k} = O(\varepsilon^{\omega(1)} n^h)$, as required. ■

2.1 Hypergraphs

The removal lemma (Theorem 1.1) was famously generalized to hypergraphs by Gowers and independently Nagle-Rödl-Schacht-Skokan in the early 2000's. Another proof was given by Tao. The statement is as follows. For k-graphs G, H, we say that G is *ε-far* from being H-free if we must delete at least εn^k edges to turn G into an H-free hypergraph, where $n = |V(G)|$.

Theorem 2.6 (Hypergraph removal lemma). *For every k-graph H and $\varepsilon > 0$, there is $\delta = \delta_H(\varepsilon) > 0$ such that if G is an n-vertex k-graph which is ε-far from being H-free, then G contains at least $\delta n^{v(H)}$ copies of H.*

The proof of 2.6 relies on an appropriate generalization of Szemerédi's regularity lemma to hypergraphs. Regularity for hypergraphs is much more involved than for graphs. We may touch upon this towards the end of the course.

It is natural to ask for an extension of Theorem 2.2 to k-graphs. Namely, for which k-graphs H does it hold that $\delta_H(\varepsilon) = \text{poly}(\varepsilon)$? A k-graph H is *k-partite* if there is a partition $V(H) = U_1 \cup \cdots \cup U_k$ such that each edge intersects each of the sets U_i. The Kővári-Sós-Turán theorem has an analogue for hypergraphs; this is a classical theorem proved by Erdős.

Theorem 2.7 (Hypergraph KST, supersaturation form). *An n-vertex k-graph with εn^k edges contains at least poly$(\varepsilon) n^{s_1 + \cdots + s_k}$ copies of $K^{(k)}_{s_1,\ldots,s_k}$.*

Theorem 2.7 implies that if H is k-partite then $\delta_H(\varepsilon) = \text{poly}(\varepsilon)$. Kohayakawa, Nagle and Rödl suggested that the converse is also true, i.e., that being k-partite is necessary for having a polynomial removal lemma. After an earlier work by Alon and Shapira, this was proved by the author and Shapira.

Theorem 2.8 (Gishboliner-Shapira 2025). *For a k-graph H, $\delta_H(\varepsilon) =$ poly(ε) if and only if H is k-partite.*

The idea for proving Theorem 2.8 is, roughly speaking, to reduce from the hypergraph case to the graph case. More precisely, we will show that for every non-k-partite H, one can reduce to a graph cycle or to a complete hypergraph. Both of these cases can be easily handled (cycles are handled as in the proof of Theorem 2.4). To explain what we mean by a reduction, we need the following definition. Let H be a k-graph and let F be an r-graph, $r \leq k$. We say that F is an *induced shadow subgraph* of H if there is a subset $X \subseteq V(H)$, $|X| = |V(F)|$, such that

1. For every $e \in E(H)$, $|e \cap X| \leq r$.

2. The r-graph $\{e \cap X : e \in E(H), |e \cap X| = r\}$ is isomorphic to F.

Very roughly speaking, our approach is to show that if F is an induced shadow subgraph of H, then hardness for the F-removal lemma implies hardness for the H-removal lemma.[8] What are the base cases of this argument? I.e., what are the hypergraphs to which we reduce all others? In the following lemma we show that it suffices to take the base cases to be (2-uniform) cycles and complete hypergraphs.

Lemma 2.9. *Let H be a non-k-partite k-graph. Then H has an induced shadow subgraph which is either a 2-uniform cycle or $K_{s+1}^{(s)}$ for some $2 \leq s \leq k$.*

Proof. Let $\partial_2 H$ be the *2-shadow* of H, i.e., $\partial_2 H$ is the graph on H where xy is an edge if there is $e \in E(H)$ with $x, y \in e$. If $\partial_2 H$ has an induced cycle C_k of length at least 4, then this gives an induced shadow copy of C_k in H.[9] Otherwise, $\partial_2 H$ is chordal. Also, $\chi(\partial_2 H) \geq k+1$, because else H is k-partite. Chordal graphs are perfect, meaning that $\omega(\partial_2 H) = \chi(\partial_2 H) \geq k+1$. Let Y be a $(k+1)$-clique in $\partial_2 H$. Let $X \subseteq Y$ be a minimal subset which is not contained in any edge of H. This is well defined because Y itself is not contained in any edge of H, as $|Y| = k+1$. Also, $|X| \geq 3$ because any subset of Y of size 2 is contained in an edge of H, as Y is a clique in $\partial_2 H$. Put $s+1 = |X|$. Then X forms an induced shadow copy of $K_{s+1}^{(s)}$ in H.[10] ∎

Let us now sketch the proof of Theorem 2.8. As alluded to above, the proof works by taking a construction for one of the base cases – i.e., a graph cycle or $K_{s+1}^{(s)}$ – and using it to get a construction for H. A construction for cycles is given by Theorem 2.4. As explained above, the cycle given by Lemma 2.9 may be even. A crucial point is that the proof of Theorem 2.4 works also for even cycles, provided that we are only interested in bounding the number of cycles of the form $(v_1, \dots, v_k, v_1)$ with $v_i \in V_i$. A construction for $K_{s+1}^{(s)}$ (i.e., showing that the $K_{s+1}^{(s)}$-removal lemma is not polynomial) is similar. For completeness, we give a sketch at the end of this section.

Proof sketch of Theorem 2.8. Let H be a non-k-partite k-graph. We assume that H is a core (we saw in the proof of Theorem 2.2 how to reduce from a hypergraph to its core via a constant-size blowup).[11] By Lemma 2.9, H has an induced shadow subgraph which is a 2-uniform cycle or $K_{s+1}^{(s)}$. We will only handle the former case; the latter case is handled similarly. So suppose that $V(H) = \{1, \dots, h\}$, and $(1, \dots, \ell, 1)$ is an induced shadow copy of C_ℓ. Use (the proof of) Theorem 2.4 to obtain a graph G' with parts $V_1, \dots, V_\ell$ having a collection $\mathcal{C}$ of εn^2 edge-disjoint cycles of the form $(v_1, \dots, v_\ell)$ but only $\varepsilon^{\omega(1)} n^\ell$ cycles of this form in total.[12] We now add additional sets $V_{\ell+1}, \dots, V_h$ and would like to define k-graph G on the basis of G'. We will extend each ℓ-cycle $(v_1, \dots, v_\ell) \in \mathcal{C}$

[8] This is not accurate. What we actually need is hardness for an F-partite version of the F-removal lemma. For example, in some cases we take F to be an even cycle. While the removal lemma for an even cycle C_ℓ *is* polynomial (by Theorem 2.2), we can in fact construct a graph with parts $V_1, \dots, V_\ell$ having εn^2 edge-disjoint copies of C_ℓ, but altogether only $\varepsilon^{\omega(1)} n^\ell$ copies of C_ℓ *which are of the form* $v_1, \dots, v_\ell$ *with* $v_i \in V_i$ (this is just the construction from the proof of Theorem 2.4). Having few such "canonical" copies suffices for our reduction (even though there are many more copies of C_ℓ of different forms, because the C_ℓ-removal lemma is polynomial).

[9] Indeed, the vertex set of this cycle C intersects each edge of H in at most two vertices, because otherwise $V(C)$ contains a triangle in $\partial_2 H$, in contradiction to C being induced of length at least 4.

[10] Indeed, by the choice of X, we have that $|e \cap X| \leq s$ for every $e \in E(H)$, and that for every $Z \subseteq X$ of size s there is $e \in E(H)$ with $e \cap X = Z$.

[11] Note that, in analogy to the graph case, if H is non-k-partite then so is its core (and if H is k-partite then its core is an edge).

[12] As explained before, if ℓ is even then G' may have (many) other cycles of length ℓ, say between two of the parts. But this will not matter for our argument.

to many (roughly n^{k-2}) copies of H in G, such that any two such copies of H either intersect in at most $k-1$ vertices, or they share at most $k-3$ vertices in $V_{\ell+1},\dots,V_h$ (and hence must share at least 3 vertices in $V_1,\dots,V_\ell$).

Claim 2.10. *There is a collection $\mathcal{H} \subseteq V_1 \times \cdots \times V_h$ with $|\mathcal{H}| \geq \Omega(\varepsilon n^k)$ such that:*

1. *For every $(v_1,\dots,v_h) \in \mathcal{H}$ it holds that $(v_1,\dots,v_\ell) \in \mathcal{C}$.*
2. *For two $H_1, H_2 \in \mathcal{H}$, if $|H_1 \cap H_2| \geq k$ then H_1, H_2 have at most $k-3$ common vertices in $V_{\ell+1},\dots,V_h$.*[13]

Proof. There are several ways of doing this: arithmetically or probabilistically. We use the probabilistic deletion method. For each $(v_1,\dots,v_\ell) \in \mathcal{C}$ and $(v_{\ell+1},\dots,v_h) \in V_{\ell+1}\times\cdots\times V_h$, add $(v_1,\dots,v_h)$ to $\mathcal{H}$ independently with probability $p = cn^{(k-2)-(h-\ell)}$, for a small constant $c > 0$. Then, for every pair H_1, H_2 violating Item 2, delete one of H_1, H_2 from $\mathcal{H}$. The expected size of $\mathcal{H}$ before the deletions is $\mathbb{E}[|\mathcal{H}|] = |\mathcal{C}|n^{h-\ell}p = c\varepsilon n^k$. Now we estimate the expected number of violations to Item 2. Suppose that H_1, H_2 is such a violation. Let $C_i = H_i \cap (V_1 \times \cdots \times V_\ell)$, so $C_i \in \mathcal{C}$. Suppose first that $C_1 = C_2$. To violate Item 2, H_1, H_2 must share at least $k-2$ vertices in $V_{\ell+1},\dots,V_h$. Hence, the number of choices for H_1, H_2 is at most $O(|\mathcal{C}|n^{2(h-\ell)-(k-2)})$. Therefore, the expected number of such pairs H_1, H_2 in $\mathcal{H}$ is at most $O(|\mathcal{C}|n^{2(h-\ell)-(k-2)}p^2) = O(c \cdot |\mathcal{C}|n^{h-\ell}p) = O(c \cdot \mathbb{E}[|\mathcal{H}|])$. Suppose now that $C_1 \neq C_2$. Any two distinct cycles in $\mathcal{C}$ share at most 1 vertex; this is a stronger property than being edge-disjoint, and it is guaranteed by the way the collection $\mathcal{C}$ is constructed in the proof of Theorem 2.4. So $|C_1 \cap C_2| \leq 1$. To violate Item 2, H_1, H_2 must share at least $k - |C_1 \cap C_2|$ vertices in $V_{\ell+1},\dots,V_h$. Suppose first that $|C_1 \cap C_2| = 0$. Then the number of choices for H_1, H_2 is at most $O(|\mathcal{C}|^2 n^{2(h-\ell)-k})$. Hence, the expected number of such pairs H_1, H_2 appearing in $\mathcal{H}$ is $O(|\mathcal{C}|^2 n^{2(h-\ell)-k}p^2) \leq O(c \cdot |\mathcal{C}|n^{h-\ell}p) = O(c \cdot \mathbb{E}[|\mathcal{H}|])$. Here we used that $|\mathcal{C}| \leq n^2$. Similarly, if $|C_1 \cap C_2| = 1$, then the number of choices for C_1, C_2 is at most $O(|\mathcal{C}|n)$ (because any two cycles in $\mathcal{C}$ share at most one vertex), and so the number of choices for H_1, H_2 is at most $O(|\mathcal{C}|n \cdot n^{2(h-\ell)-(k-1)})$. Hence, the expected number of such pairs H_1, H_2 in $\mathcal{H}$ is again $O(|\mathcal{C}|n^{2(h-\ell)-k}p^2) = O(c \cdot \mathbb{E}[|\mathcal{H}|])$. Choosing c small enough, we see that the expected number of violations is at most $\frac{1}{2}\mathbb{E}[|\mathcal{H}|]$, say, so the deletion method works. ■

Let us now continue the proof of Theorem 2.8. Define a k-graph G by putting a copy of H on each h-tuple in $\mathcal{H}$. We claim that these copies are edge-disjoint. Indeed, suppose that two such copies share an edge. Then in particular, they share at least k vertices. By Item 2 of the claim, this means that they share at most $k-3$ vertices in $V_{\ell+1},\dots,V_h$, and hence at least 3 in $V_1,\dots,V_\ell$. But this means that some edge of H intersects $\{1,\dots,\ell\}$ in at least 3 vertices, contradicting the definition of an induced shadow copy. It follows that the copies in $\mathcal{H}$ are indeed edge-disjoint. As $|\mathcal{H}| \geq \Omega(\varepsilon n^k)$, G is $\Omega(\varepsilon)$-far from H-free.

Let us now bound the total number of H-copies in G. Since H is a core and G is homomorphic to H (via the homomorphism $V_i \mapsto i$), every H-copy in G is of the form $v_1,\dots,v_h$ with $v_i \in V_i$. Consider such a copy. Fix any $1 \leq i \leq \ell$ and consider the pair v_i, v_{i+1}, with indices taken modulo ℓ. By the definition of an induced shadow copy, the pair $\{i, i+1\}$ (with indices taken modulo ℓ) is contained to some edge of H. Therefore, there is an edge $e \in E(G)$ such that $v_i, v_{i+1} \in e$. By the definition of G, e belongs to some copy $H' \in \mathcal{H}$, and by Item 1 of Claim 2.10, we have that

[13] This condition might seem a bit mysterious (although it is exactly what we need). To prove it, we will actually show that if $|H_1 \cap H_2| \geq k$ then H_1, H_2 have the same vertices in $V_1,\dots,V_\ell$ (i.e., they extend the same $(v_1,\dots,v_\ell) \in \mathcal{C}$), and that any two different extensions of $(v_1,\dots,v_\ell) \in \mathcal{C}$ share at most $k-3$ vertices in $V_{\ell+1},\dots,V_h$.

$v_i v_{i+1} \in E(G')$. Thus, for every copy H-copy $v_1, \dots, v_h$ in G, it holds that $(v_1, \dots, v_\ell)$ is an ℓ-cycle in G'. But by the choice of G', the number of such ℓ-cycles is at most $\varepsilon^{\omega(1)} n^\ell$. Hence, the number of H-copies in G is at most $\varepsilon^{\omega(1)} n^\ell \cdot n^{h-\ell} = \varepsilon^{\omega(1)} n^h$. ■

To conclude this section, let us sketch the proof of the fact that the $K_{s+1}^{(s)}$-removal lemma is not polynomial. This fact is needed for the proof of Theorem 2.8.

Theorem 2.11. *For every $s \geq 2$, there exists an n-vertex s-graph G with εn^s edge-disjoint copies of $K_{s+1}^{(s)}$, but only $\varepsilon^{\omega(1)} n^{s+1}$ copies of $K_{s+1}^{(s)}$ in total.*

Proof sketch. Take the set $S \subseteq [n]$ from Theorem 2.3, applied with parameter $k = 3$. As in the proof of Theorem 2.4, we choose n such that $|S| = \varepsilon n$. Take $s+1$ sets $V_1, \dots, V_{s+1}$ (each identified with $[(s+1)n]$), and for each $x \in S$ and $y_1, \dots, y_{s-1} \in [n]$, add a copy of $K_{s+1}^{(s)}$ on the vertices $v_1, \dots, v_{s+1}$, $v_i \in V_i$, where $v_i = y_i$ for $1 \leq i \leq s-1$, $v_s = x + \sum_{i=1}^{s-1} y_i$, $v_{s+1} = 2x + \sum_{i=1}^{s-1} y_i$. It is easy to check that these copies of $K_{s+1}^{(s)}$ are edge-disjoint, because knowing s of the vertices $v_1, \dots, v_{s+1}$ allows one to find $x, y_1, \dots, y_{s-1}$. Thus, G has a collection of $|S| n^{s-1} \geq \varepsilon n^s$ edge-disjoint copies of $K_{s+1}^{(s)}$. On the other hand, one can show, using that S has no solution to $x_1 + x_2 = 2x_3$ with distinct x_1, x_2, x_3, that every copy of $K_{s+1}^{(s)}$ in G is one of the original copies we put in. Hence, the total number of copies is $|S| n \leq n^s = n^{s+1}/n \leq \varepsilon^{\omega(1)} n^{s+1}$. ■

3 The induced removal lemma

It is natural to consider a variant of the removal lemma for induced subgraphs. In this case, we allow both adding and removing edges, since adding edges may also be useful in order to make a graph induced H-free. Thus, an n-vertex graph G is said to be *ε-far* from induced H-free if one has to add/delete at least εn^2 edges in order to make G induced H-free. The following is the induced analogue of the removal lemma:

Theorem 3.1 (Induced removal lemma, Alon-Fischer-Krivelevich-Szegedy 2000)**.** *Let H be a fixed graph. For every $\varepsilon > 0$ there is $\delta = \delta_H(\varepsilon) > 0$ such that if an n-vertex graph G is ε-far from induced H-free, then G contains at least $\delta n^{v(H)}$ induced copies of H.*

The proof of Theorem 3.1 is significantly more complicated than that of Theorem 1.1. The natural approach is to take a regular partition and try to clean it, arguing that if after the cleaning there still remains an induced copy of H, then before the cleaning there are many such copies. For the cleaning, it is natural to delete all possible edges between V_i, V_j if $d(V_i, V_j)$ is close to 0, and to add all possible edges between V_i, V_j if $d(V_i, V_j)$ is close to 1. However, it is not clear how to handle the non-regular pairs (V_i, V_j) and the edges inside the sets V_i. Whereas in the non-induced case we could simply delete all such edges and thus make sure that any remaining H-copy uses none of them, it is not clear how to proceed in the induced case. We will need a more involved "regularity scheme". Let us now describe the structure that we need in order to prove Theorem 3.1. We only present the main ideas and avoid many of the details.

Regularity scheme for induced removal

The proof of Theorem 3.1 proceeds by finding an ε-regular partition $V_1, \dots, V_t$ of G, and disjoint $U_{i,1}, \dots, U_{i,h} \subseteq V_i$, where $h = |V(H)|$ (say), such that the following holds:

1. **All** pairs $(U_{i,k}, U_{j,\ell})$ for $(i,k) \neq (j,\ell)$ are ε-regular.
2. For every $1 \leq i < j \leq t$, all pairs $(U_{i,k}, U_{j,\ell})$ for $k, \ell \in [h]$ have the same density, up to ε.
3. For every $1 \leq i \leq t$, either $d(U_{i,k}, U_{i,\ell}) \geq \frac{1}{2}$ for all $1 \leq k < \ell \leq h$, or $d(U_{i,k}, U_{i,\ell}) \leq \frac{1}{2}$ for all $1 \leq k < \ell \leq h$.
4. For all but εt^2 of the pairs $1 \leq i < j \leq t$, it holds that $|d(V_i, V_j) - d(U_{i,k}, U_{j,\ell})| \leq \varepsilon$ for all $k, \ell \in [h]$.

Cleaning the graph consists of the following:

(a) For every $1 \leq i < j \leq t$, if $d(U_{i,k}, U_{j,\ell}) \geq 1 - 2\varepsilon$ for all $k, \ell \in [h]$, then make (V_i, V_j) complete, and if $d(U_{i,k}, U_{j,\ell}) \leq 2\varepsilon$ for all $k, \ell \in [h]$, then make (V_i, V_j) empty. Else, make no changes.

(b) For every $1 \leq i \leq t$, if $d(U_{i,k}, U_{i,\ell}) \geq \frac{1}{2}$ for all $1 \leq k < \ell \leq h$ then make V_i a clique, and if $d(U_{i,k}, U_{i,\ell}) \leq \frac{1}{2}$ for all $1 \leq k < \ell \leq h$ then make V_i an independent set.

Note that in Item (a), if we make no changes between (V_i, V_j) then $\varepsilon \leq d(U_{i,k}, U_{j,\ell}) \leq 1 - \varepsilon$ for all $k, \ell \in [h]$, meaning that we can embed both edges and non-edges in these pairs $(U_{i,k}, U_{j,\ell})$.

The sets $U_{i,k}$ are sometimes called *representatives* for V_i; the changes between V_i and V_j (or within V_i) are made according to their representatives $(U_{i,k}, U_{j,\ell})$. Items 2-3 require that these representatives are consistent. Item 4 is meant to ensure that the number of changes made when cleaning the graph (in Item (a)) is small. The proof proceeds by showing that if there is a copy of H in the cleaned graph, say (without loss of generality) between some sets $V_1, \dots, V_r$ and using a_i vertices from V_i for each $i \in [r]$, then we can find many H-copies in the original graph by taking a_i representative sets $U_{i,k}$ from V_i for each $1 \leq i \leq k$.

Very roughly speaking, the structure described in Items 1-4 is found as follows: First, apply the regularity lemma to find the partition $\mathcal{P} = \{V_1, \dots, V_t\}$. Then apply the regularity lemma again with a much smaller regularity parameter (which depends on t; much smaller than $1/t$ in fact) to find a partition $\mathcal{Q}$ refining $\mathcal{P}$, and sample a set $U_i \subseteq V_i$ randomly. With high probability, all pairs (U_i, U_j) are highly regular. Now apply the regularity lemma on $G[U_i]$ (this time with parameter ε again) to partition U_i into sets $U_{i,k}$, and apply Ramsey's theorem (preceded by Turán's theorem) on these sets to find $U_{i,1}, \dots, U_{i,h}$ satisfying Item 3. Items 1-2 hold because all pairs (U_i, U_j) are highly regular.

To satisfy Item 4, more is required, and in fact the above presentation is somewhat misleading: One applies the regularity lemma not just twice but repeatedly, obtaining a sequence of partitions $\mathcal{P}_i$ such that $\mathcal{P}_{i+1}$ refines $\mathcal{P}_i$ and is regular with a parameter appropriately defined in terms of $|\mathcal{P}_i|$. One stops when $q(\mathcal{P}_{i+1}) \leq q(\mathcal{P}_i) + \varepsilon$, where $q(\cdot)$ is the mean square density. It is then possible to show that Item 4 holds. This argument proves the so-called *strong regularity lemma*, which we shall not go into. In the following section, we will see in more detail a variant of the above regularity scheme.

The infinite removal lemma

It is natural to ask for an analogue of the removal lemma for families of forbidden (induced) subgraphs. The more general result of this type was obtained by Alon and Shapira, and applies to *any* (possibly infinite) graph family.

Theorem 3.2 (Infinite removal lemma, Alon-Shapira 2005). *Let $\mathcal{H}$ be a (possibly infinite) family of graphs. For every $\varepsilon > 0$ there exist $\delta = \delta_{\mathcal{H}}(\varepsilon) > 0$ and $m = m_{\mathcal{H}}(\varepsilon) \geq 1$ such that if an n-vertex graph G is ε-far from induced $\mathcal{H}$-freeness, then there is $H \in \mathcal{H}$ with $|V(H)| \leq m$, such that G contains at least $\delta n^{v(H)}$ copies of H.*

Polynomial removal lemmas

A key question is to characterize the graph-families $\mathcal{H}$ for which $\delta_{\mathcal{H}}(\varepsilon)$ and $m_{\mathcal{H}}(\varepsilon)$ in Theorem 3.2 depend polynomially on ε (more precisely, on ε and $1/\varepsilon$, respectively).

Problem 3.3. *For which graph-families $\mathcal{H}$ is the induced $\mathcal{H}$-free removal lemma polynomial?*

A special case of this problem (for the property of not-necessarily-induced H-freeness) was handled in Section 2. Returning to the induced H-removal lemma for a single graph H, works of Alon-Shapira and Alon-Fox give the following almost complete characterization. We use P_k (resp. C_k) to denote the path (resp. cycle) with k vertices.

Theorem 3.4 (Alon-Shapira 2006, Alon-Fox 2015)**.**

1. *If $H \in \{P_2, \overline{P_2}, P_3, \overline{P_3}, P_4\}$ then the induced H-removal lemma is polynomial.*
2. *If $H \notin \{P_2, \overline{P_2}, P_3, \overline{P_3}, P_4, C_4, \overline{C_4}\}$ then the induced-H removal lemma is not polynomial.*

The only remaining case is $H = C_4$.[14] This case remains open, but an exponential bound is known:

Theorem 3.5 (Gishboliner-Shapira 2019)**.** *For the induced-C_4 removal lemma, we have*

$$\delta_{C_4}(\varepsilon) \geq 2^{-poly(1/\varepsilon)}.$$

Conjecture 3.6. *The induced-C_4 removal lemma is polynomial.*

4 VC-dimension and ultra-strong regularity

We begin by recalling the definition of VC-dimension, and then discuss its connection to graphs and regularity.

Definition 4.1 (Shattered set, VC-dimension)**.** *Let $\mathcal{F}$ be a family of subsets of a set V. A set $S \subseteq V$ is* shattered *by $\mathcal{F}$ if for every $T \subseteq S$, there exists $F \in \mathcal{F}$ with $S \cap F = T$. The* VC-dimension *of $\mathcal{F}$ is the maximum size of a shattered set.*

VC-dimension is a fundamental measure of complexity used in combinatorics and computer science. One of the basic facts about VC-dimension is the so-called Sauer-Shelah lemma, stating the following:

Theorem 4.2 (Sauer-Shelah lemma)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of subsets of $[n]$ with VC-dimension d. Then $|\mathcal{F}| \leq \sum_{i=0}^{d} \binom{n}{i}$.*

Note that the bound in Theorem 4.2 is tight, because the set family consisting of all sets of size at most d has VC dimension d.

Proof of Theorem 4.2. We prove by induction on n that the number of sets shattered by $\mathcal{F}$ is at least $|\mathcal{F}|$. This suffices, because if $|\mathcal{F}| > \sum_{i=0}^{d} \binom{n}{i}$ then there must exist a shattered set of size larger than $d+1$, and hence the VC-dimension is larger than $d+1$.

Write $\mathcal{F}_0 = \{F \in \mathcal{F} : n \notin F\}$ and $\mathcal{F}_1 = \{F \setminus \{n\} : F \in \mathcal{F}, n \in F\}$. Then $\mathcal{F}_0, \mathcal{F}_1 \subseteq 2^{[n-1]}$, and $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$. By induction, $\mathcal{F}_i$ shatters at least $|\mathcal{F}_i|$ sets for $i = 0, 1$. Every set shattered by $\mathcal{F}_0$ or

[14]The case H and $\overline{H}$ are equivalent.

$\mathcal{F}_1$ is (trivially) shattered by $\mathcal{F}$. Also, it is easy to see that if S is shattered by both $\mathcal{F}_0, \mathcal{F}_1$, then $S \cup \{n\}$ is shattered by $\mathcal{F}$. This allows us to conclude that

$$\#\{\text{sets shattered by } \mathcal{F}\} \geq \#\{\text{sets shattered by } \mathcal{F}_0\} + \#\{\text{sets shattered by } \mathcal{F}_1\} \geq |\mathcal{F}_0| + |\mathcal{F}_1| = |\mathcal{F}|.$$

■

In applications, the exact bound $\sum_{i=0}^{d} \binom{n}{i}$ from Theorem 4.2 is often not important; the crucial fact is that $|\mathcal{F}|$ is polynomial in n.

To use VC-dimension in the context of graphs, we consider the following set-family: Let G be a graph, and let $\mathcal{F} = \{N(v) : v \in V(G)\}$. The VC-dimension of G is defined as the VC-dimension of the set-family $\mathcal{F}$. What does it mean for a graph to have unbounded VC-dimension? It means that for every fixed $d \geq 1$, there is a set $S = \{x_1, \dots, x_d\}$ shattered by $\mathcal{F}$. This in turn means that there are vertices $(y_I : I \subseteq [d])$, such that y_I is adjacent to x_i if and only if $i \in I$. In what follows, we will want to assume that $y_I \notin \{x_1, \dots, x_d\}$ for every I. This can be achieved by taking a slightly larger shattered set $S = \{x_1, \dots, x_d, x_{d+1}, \dots, x_{d+k}\}$, where we think of $x_{d+1}, \dots, x_{d+k}$ as "dummy vertices". Doing this supplies us with 2^k different vertices to play the role of y_I for each $I \subseteq [d]$, so if $2^k > d$ then we can choose such a y_I which is outside $\{x_1, \dots, x_d\}$.

We now see that if G has unbounded VC-dimension, then it has a bi-induced copy of every fixed-size bipartite graph $H = (A, B)$ (recall Definition 1.6). Indeed, we can construct such a copy by first choosing a shattered set $S = \{x_1, \dots, x_d\}$ to play the role of A, and then, for each $b \in B$, choosing y_I for I which corresponds to the neighborhood of b in A (namely, if $A = \{a_1, \dots, a_d\}$, then I is the set of all $i \in [d]$ such that $a_i b \in E(H)$). By taking a slightly larger set S (of size $d + k$ for $2^k \geq |B|$, as above), we can make sure that we have enough vertices y_I with neighborhood $\{x_i : i \in I\}$ in $\{x_1, \dots, x_d\}$, in case several vertices in B have the same neighborhood in H. We thus conclude the following:

Fact 4.3. *G has unbounded VC-dimension if and only if G contains a bi-induced copies of all fixed-size bipartite graphs.*[15]

The above is of course not a rigorous statement (because of the term "unbounded"), but it should be clear what it means. A rigorous statement would be that if G has VC-dimension at least d_1, then it contains bi-induced copies of all bipartite graphs of size d_2 (for some d_2 growing with d_1), and vice versa.

What can we say about a graph G which avoids bi-induced copies of some fixed bipartite H? Let us apply the regularity lemma to obtain an ε-regular equipartition $V_1, \dots, V_t$. By Lemma 1.7, all regular pairs (V_i, V_j) have density at most γ or at least $1 - \gamma$, provided that $\varepsilon \ll \gamma$. Thus, there is an equipartition of $V(G)$ where all but γt^2 of the pairs (V_i, V_j) have density at most γ or at least $1 - \gamma$.[16] Such a partition is called *γ-homogeneous*. Thus, bounded VC-dimension implies the existence of γ-homogeneous partitions. However, the partition obtain in this way is very large, of tower-type size in $1/\gamma$. Can we do better? As we will now show, using the Sauer-Shelah lemma we can find a partition of size only polynomial in $1/\gamma$. The key fact we will need is as follows:

Lemma 4.4. *If G has VC-dimension d, then for every $\varepsilon > 0$, there are vertices $x_1, \dots, x_t$, $t \leq (1/\varepsilon)^{O(d)}$, such that for every $x \in V(G)$ there is $i \in [t]$ with $|N(x) \triangle N(x_i)| \leq \varepsilon n$.*

[15] We only proved one direction, but the other direction is also easy: Take the $d \times 2^d$ incidence bipartite graph of $[d]$ versus subsets of $[d]$. If G contains a bi-induced copy of this bipartite graph, then its VC-dimension is at least d.

[16] Such a pair is called *γ-homogeneous*. Note that a being homogeneous is a stronger property than being regular: a γ-homogeneous pair is necessarily $\gamma^{1/3}$-regular. This is left as an exercise for the reader.

Proof. Let $x_1, \dots, x_t$ be a maximum collection of elements such that $|N(x_i) \triangle N(x_j)| > \varepsilon n$ for every $1 \le i < j \le t$. It suffices to show that $t < t_0 := (1/\varepsilon)^{d+1}$.[17] Suppose not, and suppose that $t = t_0$ (by disposing of the other x_i's). Sample a subset $U \subseteq V(G)$ of size $|U| = m := \frac{2\log(t)}{\varepsilon} = \tilde{O}(\frac{1}{\varepsilon})$. For $1 \le i < j \le t$, the probability that $N(x_i) \cap U = N(x_j) \cap U$ is at most

$$\frac{\binom{(1-\varepsilon)n}{m}}{\binom{n}{m}} \le (1-\varepsilon)^m \le e^{-\varepsilon m} = t^{-2}.$$

By the union bound, there is an outcome for U such that $N(x_i) \cap U \ne N(x_j) \cap U$ for every $1 \le i < j \le t$. But now $(N(x_i) \cap U : 1 \le i \le t)$ is a set system on U of size $t = (1/\varepsilon)^{d+1} > \sum_{i=0}^{d} \binom{|U|}{i}$, so by the Sauer-Shelah lemma (Theorem 4.2), it has VC-dimension larger than d, in contradiction to the assumption that G has VC-dimension d. ■

We will now use Lemma 4.4 to find a small ε-homogeneous equipartition of a graph G with bounded VC-dimension. As far as we know, this result is originally due to Lovász-Szegedy and Alon-Fischer-Newman.

Theorem 4.5 (Lovász-Szegedy 2010, Alon-Fischer-Newman 2007)**.** *If G has VC-dimension d, then it has an ε-homogeneous equipartition of size $(1/\varepsilon)^{O(d)}$.*

Proof. Let $x_1, \dots, x_t \in V(G)$, $t \le (1/\gamma)^{O(d)}$, be the vertices given by Lemma 4.4, applied with parameter $\gamma = \text{poly}(\varepsilon) \ll \varepsilon$ to be chosen (implicitly) later. For each $i \in [t]$, let X_i be the set of all $x \in V(G)$ such that $|N(x) \triangle N(x_i)| \le \gamma n$. So $X_1 \cup \dots \cup X_t = V(G)$ by the guarantees of Lemma 4.4. The idea is to claim that $X_1, \dots, X_t$ is an ε-homogeneous partition. The partition $X_1, \dots, X_t$ is not an equipartition, so we need to adapt the definition of an ε-homogeneous partition to allow parts of different sizes: A partition $X_1, \dots, X_t$ is ε-homogeneous if the sum of $|X_i||X_j|$ over all pairs $1 \le i, j \le t$ with $\varepsilon < d(X_i, X_j) < 1 - \varepsilon$ is at most εn^2. This sum includes terms $i = j$.

Let us now show that $\{X_1, \dots, X_t\}$ is ε-homogeneous. Sample vertices $x, y \in V(G)$ uniformly at random and then a vertex x' belonging to the same part X_i as x. Let $\mathcal{A}$ be the event that $xy \in E(G)$ but $x'y \notin E(G)$, or vice versa. In other words, this is the event that $y \in N(x) \triangle N(x')$. We will show that due to the choice of $X_1, \dots, X_t$, $\mathbb{P}[\mathcal{A}] \le 2\gamma$, and that this implies that $\{X_1, \dots, X_t\}$ is ε-homogeneous. First, condition on the choice of x, x'. Since x, x' are in the same part X_i, we have $|N(x) \triangle N(x')| \le |N(x) \triangle N(x_i)| + |N(x') \triangle N(x_i)| \le 2\gamma n$ (by the triangle inequality), so $\mathbb{P}[y \in N(x) \triangle N(x')] \le 2\gamma$. It follows that $\mathbb{P}[\mathcal{A}] \le 2\gamma$.

Now suppose by contradiction that $\{X_1, \dots, X_t\}$ is not ε-homogeneous. Fix a pair X_i, X_j with $\varepsilon < d(X_i, X_j) < 1 - \varepsilon$. We need the following claim:

Claim 4.6. *For disjoint vertex-sets U, V, if $\varepsilon < d(U, V) < 1 - \varepsilon$, then there are $\Omega(\varepsilon)|U|^2|V|$ triples $x, x' \in U, y \in V$ or $\Omega(\varepsilon)|V|^2|U|$ triples $x, x' \in V, y \in U$ satisfying $y \in N(x) \triangle N(x')$.*

The claim is left as an exercise for the reader.

By the claim, without loss of generality there are $\Omega(\varepsilon)|X_i|^2|X_j|$ triples $x, x' \in X_i, y \in X_j$ with $y \in N(x) \triangle N(x')$. The probability that the random vertices x, x', y form such a triple is

$$\frac{\Omega(\varepsilon)|X_i|^2|X_j|}{n^2|X_i|} = \frac{\Omega(\varepsilon)|X_i||X_j|}{n^2}.$$

[17] As one can see from the proof, $(1/\varepsilon)^{d+1}$ can be replaced with $\tilde{O}(1/\varepsilon^d)$.

Summing over all non-ε-homogeneous pairs (X_i, X_j) and using the assumption that $\{X_1, \dots, X_t\}$ is not ε-homogeneous, we get $\mathbb{P}[\mathcal{A}] \geq \Omega(\varepsilon^2) > 2\gamma$, provided that γ is small enough. This is a contradiction.

A last step, in order to obtain an equipartition, is to chop up each part X_i into equal-sized parts plus maybe one leftover part, then collect the leftover parts and partition them again into equal-sized parts. One can show that (if the part size is small enough) then the resulting partition is still β-homogeneous for β which depends polynomially on ε. (A refinement of an ε-homogeneous partition is $O(\sqrt{\varepsilon})$-homogeneous, which follows from Markov's inequality). ■

Alon, Fischer and Newman in fact proved a stronger statement: They show that it suffices to assume that G has few bi-induced copies of some fixed bipartite graph H (instead of assuming that G has no such copies at all). This result, as well as Theorem 4.5, are sometimes called an *ultra-strong regularity lemma*.

Theorem 4.7 (Alon-Fischer-Newman 2007)**.** *For every bipartite graph H and $\varepsilon > 0$, there is $\delta = poly(\varepsilon) > 0$, such that the following holds. If an n-vertex graph G has at most $\delta n^{v(H)}$ bi-induced copies of H, then G has an ε-homogeneous equipartition into at most $\frac{1}{\delta}$ parts.*

Let us now translate the condition of avoiding bi-induced copies of a bipartite graph to the condition of avoiding induced subgraphs of certain types. A *co-bipartite* graph is the complement of a bipartite graph. A *split* graph is a graph whose vertex-set can be partitioned into a clique and an independent set.

Lemma 4.8. *Let F_1 be a bipartite graph, F_2 be a co-bipartite graph, F_3 be a split graph. There is a bipartite graph $H = (A, B)$ such that if G has no induced copies of F_1, F_2, F_3, then it has no bi-induced copies of H.*

Note that the three graph types (bipartite, co-bipartite, split) capture all possibilities of partitioning the vertex-set into two homogeneous sets (cliques or independent sets).

Proof sketch of Lemma 4.8. We need to show that there is a bipartite $H = (A, B)$ such that no matter how we place edges inside A and inside B, we get a graph which contains an induced copy of F_1, F_2 or F_3. We show that a large enough random graph H satisfies this. Let the edges inside A and B be given. We can use Ramsey's theorem to partition almost all of A and of B into homogeneous sets of size k, where $k \geq |V(F_i)|$ for $i = 1, 2, 3$. Now, for such a partition, the probability that there is no induced copy of F_1, F_2, F_3 is at most $1 - 2^{-k^2}$. Thus, if $|A| = |B| = m$, the probability of failure is at most $(1 - 2^{-k^2})^{m^2/k^2} \leq e^{-\Omega_k(m^2)}$. On the other hand, the number of partitions is at most m^{2m}, so we can take a union bound. ■

By combining Theorem 4.7 and Lemma 4.8, one can prove the following:

Theorem 4.9 (Gishboliner-Shapira 2017)**.** *If $\mathcal{H}$ is a finite graph-family containing a bipartite graph, a co-bipartite graph and a split graph, then the induced-$\mathcal{H}$ removal lemma has polynomial bounds.*

Proof sketch. The proof follows the scheme described in Section 3. Let $F_1, F_2, F_3 \in \mathcal{H}$ such that F_1 is bipartite, F_2 is co-bipartite and F_3 is split. Let H be the bipartite graph given by Lemma 4.8. Let G be an n-vertex graph which is ε-far from being induced-$\mathcal{H}$-free. Suppose first that G contains at least $\delta n^{v(H)}$ bi-induced copies of H.[18] Then by Lemma 4.8, there is $i = 1, 2, 3$ such that

[18]For the sake of keeping the presentation simple, we will not choose δ explicitly. Rather, δ is the number given by Theorem 4.7; we will apply this theorem several times with different parameters, and δ is the minimum of the resulting numbers.

G contains at least $\frac{\delta}{3}n^{v(H)}$ vertex-sets of size $v(H)$ which contain an induced copy of F_i. On the other hand, each induced copy of F_i is in at most $n^{v(H)-v(F_i)}$ such vertex sets, so there are at least $\frac{\delta}{3}n^{v(F_i)}$ induced copies of F_i, as required.

From now on, suppose that G has less than $\delta n^{v(H)}$ bi-induced copies of H. We apply Theorem 4.7 to get an ε-homogeneous equipartition $\mathcal{P} = \{V_1, \dots, V_t\}$, and then apply Theorem 4.7 again to find an ε'-homogeneous equipartition Q which refines[19] $\mathcal{P}$, where ε' is small enough (but still polynomial) in terms of ε and t; i.e., $\varepsilon' = (\varepsilon/t)^C$ for a large constant C (depending on $\mathcal{H}$). Then, for each $1 \leq i \leq t$, sample $U_i \in \mathcal{Q}$ uniformly at random among all parts of $\mathcal{Q}$ which are contained in V_i. One can show that due to the choice of ε', the following holds with positive probability:

(i) For every $1 \leq i < j \leq t$, (U_i, U_j) is ε'-homogeneous; i.e., $d(U_i, U_j) \leq \varepsilon'$ or $d(U_i, U_j) \geq 1 - \varepsilon'$.

(ii) For all but $\sqrt{\varepsilon}t^2$ of the pairs $1 \leq i < j \leq t$, it holds that $|d(U_i, U_j) - d(V_i, V_j)| \leq 10\sqrt{\varepsilon}$.

The next step is as follows: For each $i \in [t]$, apply Theorem 4.7 to $G[U_i]$ to get an ε-homogeneous equipartition $\mathcal{R}_i$ of U_i. Recall that this means that all but $\varepsilon|\mathcal{R}_i|^2$ of the pairs of parts in $\mathcal{R}_i$ are ε-homogeneous (have density at most ε or at least $1 - \varepsilon$). Apply Turán's theorem to pass to $\mathcal{R}'_i \subseteq \mathcal{R}_i$ of size roughly $|\mathcal{R}'_i| \approx \frac{1}{\varepsilon}$ such that any two parts in $\mathcal{R}'_i$ are ε-homogeneous, and then apply Ramsey's theorem to find $\mathcal{R}''_i = \{U_{i,1}, \dots, U_{i,h}\} \subseteq \mathcal{R}'_i$ such that either $d(U_{i,k}, U_{i,\ell}) \leq \varepsilon$ for all $1 \leq k < \ell \leq h$ or $d(U_{i,k}, U_{i,\ell}) \geq 1 - \varepsilon$ for all $1 \leq k < \ell \leq h$.[20] Another important point is that since (U_i, U_j) is ε'-homogeneous and ε' is very small (but still polynomial) compared to ε, we have $|d(U_{i,k}, U_{j,\ell}) - d(U_i, U_j)| \leq \varepsilon$ for all $1 \leq i < j \leq t$ and $k, \ell \in [h]$.

We now achieved the setting described by Items 1-4 in Section 3. Now clean the graph as described in that section. One then shows that if the cleaned graph has an induced copy of some $F \in \mathcal{H}$, then the original graph has many (i.e., $\delta n^{v(F)}$) such induced copies. ■

Gishboliner and Shapira also proved that if a finite $\mathcal{H}$ contains no bipartite graph or no co-bipartite graph, then the induced-$\mathcal{H}$ removal lemma is not polynomial (this proof uses similar constructions to those used in Section 2). The following remains open:

Problem 4.10. *Characterize the finite graph-families $\mathcal{H}$ for which the induced-$\mathcal{H}$ removal lemma is polynomial.*

This is of course a special case of Problem 3.3. The first open case is again $\mathcal{H} = \{C_4\}$. Note that C_4 is both bipartite and co-bipartite, but not split, so the aforementioned results of Gishboliner and Shapira do not apply.

5 Property testing

Let us consider the following equivalent form of Theorem 3.2:

Theorem 5.1 (Infinite removal lemma, sampling formulation)**.** *Let $\mathcal{H}$ be a family of graphs. For every $\varepsilon > 0$ there is an integer $q = q_{\mathcal{H}}(\varepsilon)$ such that if G is ε-far from induced $\mathcal{H}$-free, then with probability at least 0.99, a sample of q vertices from G is not induced $\mathcal{H}$-free.*

[19]While this is not part of the statement of Theorem 4.7, the theorem can be reproved to allow for a partition $\mathcal{P}$ as part of the input, such that the outputted equipartition refines $\mathcal{P}$.

[20]Here h can be chosen as $h = \max_{F \in \mathcal{F}} v(F)$.

To see that the above follows from Theorem 3.2, note that a sample of $m = m_{\mathcal{H}}(\varepsilon)$ vertices contains an induced copy of $H \in \mathcal{H}$ with probability at least $\delta = \delta_{\mathcal{H}}(\varepsilon)$, so a sample of $q = Cm/\delta$ contains such a copy with probability tending to 1 as C tends to infinity. The reverse direction is also true, i.e., that Theorem 5.1 implies Theorem 3.2 (this is left as an exercise for the reader), and q depends polynomially on $m, 1/\delta$.

Theorem 5.1 leads to the notion of *property testing.* A *property tester* for a graph property $\mathcal{P}$ is a randomized algorithm that distinguishes between graphs which satisfy $\mathcal{P}$ and graphs that are ε-far from $\mathcal{P}$, with success probability at least 0.99 (say) in both cases. Namely, if an input G satisfies $\mathcal{P}$ then the algorithm must accept G with probability at least 0.99, and if G is ε-far from $\mathcal{P}$ then the algorithm must reject $\mathcal{P}$ with probability at least 0.99.[21] The algorithm works by sampling vertices and making edge queries, i.e., asking if a pair of vertices u, v forms an edge. We require that the *sample complexity* of the algorithm, i.e., the number of vertices it samples, depends only on ε and not on the size of the input graph G.

Property testing originated in the 1990s, and has since been thoroughly studied. The model we discuss here is called the *dense graph model.* There are also other models of property testing, e.g., for constant-degree graphs.

If $\mathcal{P}$ is *hereditary*, i.e., closed under the removal of vertices[22], then then there is a very simple tester for $\mathcal{P}$: Simply sample q vertices of the input graph G, and accept if and only if the subgraph induced by the sample satisfies $\mathcal{P}$. As the property is hereditary, if G satisfies $\mathcal{P}$ then the tester accepts with probability 1.[23] The fact that this algorithm is correct is simply the statement of Theorem 5.1.

One of the early and highly influential works on property testing is a paper of Goldreich, Goldwasser and Ron, where several natural graph properties were shown to be testable with polynomial sample complexity. Two key examples are k-colorability and having an independent set of size at least ρn (for a fixed $\rho \in [0, 1]$). Note that the latter is not a hereditary property. To illustrate some of the ideas in this work, let us show that bipartiteness is testable with sample complexity poly$(1/\varepsilon)$.

Theorem 5.2 (Goldreich-Goldwasser-Ron 1998)**.** *If G is ε-far from being bipartite then a sample of $q = \tilde{O}(1/\varepsilon^2)$ vertices of G induces a non-bipartite graph with probability at least* 0.9*.*

Proof. First, by deleting at most $\frac{\varepsilon}{2}n^2$ edges, we may pass to a (spanning) subgraph of G where every vertex has degree 0 or at least $\frac{\varepsilon}{2}n$. Indeed, as long as there is a vertex v with degree at least 1 but less than $\frac{\varepsilon}{2}n$, delete all edges touching v. Note that the remaining graph (after the edge deletions) is $\frac{\varepsilon}{2}$-far from bipartiteness (because G is ε-far from bipartiteness). Let U be the set of vertices which are not isolated.

The key idea is to sample in two stages. First, sample vertices $x_1, \dots, x_s$, $s = \frac{2}{\varepsilon}\log(\frac{100}{\varepsilon})$. For a vertex $u \in U$, the probability that u has no neighbors in $X = \{x_1, \dots, x_s\}$ is at most

$$\left(1 - \frac{\varepsilon}{2}\right)^s \leq e^{-\varepsilon s/2} \leq \frac{\varepsilon}{100}.$$

Let U' be the set of $u \in U$ which have a neighbor in X. By the above and Markov's inequality, with probability at least 0.95 we have $|U'| \geq |U| - \frac{\varepsilon}{5}n$. Make the vertices in $U \setminus U'$ isolated by deleting at most additional $\frac{\varepsilon}{5}n^2$ edges.

[21] In the intermediate range – i.e. that G doesn't satisfy $\mathcal{P}$ but is ε-close to it – there is no requirement on the algorithm.

[22] Note that a graph property is hereditary if and only if it is defined in terms of a (possibly infinite) family of forbidden induced subgraphs.

[23] Accepting inputs satisfying $\mathcal{P}$ with probability 1 is known as having *one-sided error.*

If $G[X]$ is not bipartite then we are already done. Otherwise, fix any bipartition (A_1, A_2) of $G[X]$. For $i = 1, 2$, let U_i be the set of all vertices $u \in U'$ which have a neighbor in A_i (if u has a neighbor in both A_1, A_2 then place u in one of the sets U_1, U_2 arbitrarily). Then $U' = U_1 \cup U_2$. Also, every vertex of G which is not currently isolated belongs to $U_1 \cup U_2$. As the remaining graph is $\frac{\varepsilon}{4}$-far from bipartiteness, there are at least $\frac{\varepsilon}{4}n^2$ edges which are inside U_1 or inside U_2. Now sample additional vertices $Y = \{y_1, \dots, y_t\}$, $t = \frac{100s}{\varepsilon} = \tilde{O}(1/\varepsilon^2)$. Note that if we sample any of the edges inside U_1 or U_2, then the bipartition (A_1, A_2) of $G[X]$ cannot be extended to a bipartition of $G[X \cup Y]$. The probability that we sample no such edge is at most

$$\left(1 - \frac{\varepsilon}{4}\right)^{t/2} \leq e^{-\varepsilon t/8} < 0.05 \cdot 2^{-s}.$$

By taking a union bound over all at most 2^s bipartitions (A_1, A_2) of $G[X]$, we see that the probability that $G[X \cup Y]$ is bipartite is at most 0.1, as required. ■

We note that the bound on q in Theorem 5.2 has been improved to $\tilde{O}(1/\varepsilon)$, which is optimal up to the logarithmic terms. Also, such a bound has been proved for much more general testing tasks, such as testing hypergraph k-colorability and (more generally) testing satisfiability.

We now move on to testing for independent sets; more precisely, for the property of containing an independent set of size at least ρn. As mentioned above, this property was shown to be testable already by Goldreich, Goldwasser and Ron. However, very recently, a new proof was discovered by Blais and Seth, which uses the container method and supplies optimal bounds on the sample complexity of such a tester. Here we present a version of their argument with somewhat weaker bounds, for the sake of simplicity.

Theorem 5.3 (Blais-Seth 2023)**.** *Let G be an n-vertex graph which is ε-far from containing an independent set of size at least ρn. Then with probability at least 0.9, a sample $X = \{x_1, \dots, x_q\}$ of $q = \tilde{O}(1/\varepsilon^3)$ vertices from G satisfies that $\alpha(G[X]) \leq (\rho - \frac{\varepsilon}{4})q$.*

Proof. The assumption on G implies that every vertex-set $U \subseteq V(G)$ of size at least ρn contains at least εn^2 edges. But in fact we can get a bit more: every vertex set U of size at least $(\rho - \frac{\varepsilon}{2})n$ contains at least $\frac{\varepsilon}{2}n^2$ edges. Indeed, otherwise, add arbitrary $\frac{\varepsilon}{2}n$ vertices to U, delete all (at most $\frac{\varepsilon}{2}n^2$ edges) touching these vertices, and delete all edges inside U. This gives an independent set of size at least ρn, in contradiction to our assumption.

The key part of the argument is the following claim, which uses the container algorithm:

Claim 5.4. *Set $t := \frac{1}{\varepsilon}$. For every independent set $I \subseteq V(G)$ with $|I| > t$, there are sets $F = F(I) \subseteq I$ and $C = C(I) \subseteq V(G)$ such that $I \subseteq F \cup C$, $|F| \leq t$ and $|C| \leq (\rho - \frac{\varepsilon}{2})n$. Furthermore, C depends only on F (and not on I). More precisely, if $F = F(I)$, then $C(I) = C(F)$.*[24]

Proof. The container algorithm is as follows: Initialize $F_0 = \emptyset$ and $C_0 = V(G)$. For $i \geq 0$, if $|C_i| \leq (\rho - \frac{\varepsilon}{2})n$ or $C_i \cap I = \emptyset$, then stop and output $F := F_i, C := C_i$. Otherwise proceed as follows:

1. Order the elements of C_i as $v_1, \dots, v_m$, such that for every $1 \leq j \leq m$, v_j has maximum degree in $G[\{v_j, \dots, v_m\}]$.[25]

2. Let $1 \leq j \leq m$ be minimal with $v_j \in I$.

[24] The set F is usually called the *fingerprint* of I, and the set C is called the *container* corresponding to F.
[25] If two vertices have the same degree, then ties are broken according to a pre-fixed ordering on $V(G)$.

3. Move v_j to F, and delete from C_i the vertices $v_1, \dots, v_j$ and all neighbors of v_j in $\{v_{j+1}, \dots, v_m\}$. Namely, set

$$F_{i+1} = F_i \cup \{v_j\}$$
$$C_{i+1} = C_i \setminus \{v_k : k \le j \text{ or } v_j v_k \in E\}.$$

It is easy to see that $F_i \subseteq I$ and $I \subseteq F_i \cup C_i$ throughout the process, and the upper bound $|C| \le (\rho - \frac{\varepsilon}{2})n$ is guaranteed by the process. The fact that C depends only on F is a standard fact about the container algorithm, and is left for the reader. It now suffices to show that the process stops in at most t steps, as this would guarantee the upper bound on F. Suppose otherwise. Consider any step i in the process except the last. Since the process did not stop at step $i+1$, we have $|C_{i+1}| \ge (\rho - \frac{\varepsilon}{2})n$. Considering the ordering $v_1, \dots, v_m$ at step i (see Item 1 above), we have $C_{i+1} \subseteq \{v_{j+1}, \dots, v_m\}$, so in particular, the set $U := \{v_j, \dots, v_m\}$ has size at least $(\rho - \frac{\varepsilon}{2})n$. As explained above, this means that U contains at least $\frac{\varepsilon}{2}n^2$ edges. As v_j is chosen as a vertex of maximum degree in $G[U]$, we have $d_U(v_j) \ge \varepsilon n$. Hence, at least εn vertices are removed from C_i at this step. As this is true for every step except the last, and we assumed that the process lasts at least t steps, we get $|C_{t-1}| \le n - (t-1) \cdot \varepsilon n < (\rho - \frac{\varepsilon}{2})n$, a contradiction to the assumption that the process did not stop before step t. ■

Let us now prove Theorem 5.3. Sample vertices $X = \{x_1, \dots, x_q\}$ uniformly at random and independently.[26] We need to upper-bound the probability that $G[X]$ contains an independent set of size at least $(\rho - \frac{\varepsilon}{4})q$. Fix any such set I, and let $F = F(I)$ and $C = C(F)$ be given by Claim 5.4. If $I \subseteq X$, then $F \subseteq X$ and X contains at least $|I| - |F|$ vertices of C. Note that $|I| - |F| \ge (\rho - \frac{\varepsilon}{4})q - \frac{1}{\varepsilon} \ge (\rho - \frac{\varepsilon}{3})q$ (provided that ε is small enough). Thus, we see that if $\alpha(G[X]) \ge (\rho - \frac{\varepsilon}{4})q$, then there is a set $F \subseteq X = \{x_1, \dots, x_q\}$ such that $|X \cap C(F)| \ge (\rho - \frac{\varepsilon}{3})q$. We union bound over the choice of indices in $[q]$ which play the role of F, and then condition on the outcome of these indices (i.e., we condition on F). The number of choices for the index set is $\binom{q}{\le t} \le e^{O(\frac{1}{\varepsilon}\log q)}$. Having conditioned on F and setting $C = C(F)$, the random variable $|X \cap C|$ is distributed as $\text{Bin}(q, |C|/n)$, and so has expectation at most $(\rho - \frac{\varepsilon}{2})q$. By Hoeffding's inequality, the probability that $|X \cap C| \ge (\rho - \frac{\varepsilon}{3})q$ is at most $e^{-\Omega(q\varepsilon^2)}$. Combining the above, we see that

$$\mathbb{P}\left[\alpha(G[X]) \ge \left(\rho - \frac{\varepsilon}{4}\right) q\right] \le e^{O(\frac{1}{\varepsilon}\log q)} \cdot e^{-\Omega(q\varepsilon^2)},$$

which is less than 0.1 if $q = \frac{C}{\varepsilon^3}\log\frac{1}{\varepsilon}$ for a large enough constant C. ■

Using Theorem 5.3, we can now obtain a tester for large independent sets.

Theorem 5.5 (Golreich-Goldwasser-Ron 1998, Blais-Seth 2023 (optimal bound))**.** *The property of containing an independent set of size at least ρn is testable with sample complexity poly$(1/\varepsilon)$.*

Proof. The algorithm samples vertices $X = \{x_1, \dots, x_q\}$, where $q = \tilde{O}(1/\varepsilon^3)$ is given by Theorem 5.3, and accepts if and only if $\alpha(G[X]) > (\rho - \frac{\varepsilon}{4})q$. If G has an independent set I of size at least ρn, then one can show using Hoeffding's inequality that $|X \cap I| > (\rho - \frac{\varepsilon}{4})q$ with probability at least 0.9. And if G is ε-far from containing an independent set of size at least ρn, then $\alpha(G[X]) \le (\rho - \frac{\varepsilon}{4})q$ with probability at least 0.9 by Theorem 5.3. ■

[26] Here we sample with repetition for the sake of simplicity. Alternatively, one can sample a subset $X \subseteq V(G)$ of size q uniformly at random and use concentration inequalities for the hypergeometric distribution.

6 The induced-C_4 removal lemma

In this section we give a sketch of the proof of Theorem 3.5, which gives an exponential bound for the induced-C_4 removal lemma. One of the key facts about induced-C_4-free graphs that we use is the following:

Theorem 6.1 (Gyárfás-Hubenko-Solymosi 2002)**.** *If an n-vertex induced-C_4-free graph has at least βn^2 edges, then it has a clique of size at least $\frac{\alpha^2}{8}n$.*

Proof. The graph G has average degree at least $2\beta n$, so it has a subgraph with minimum degree at least βn. With a slight abuse of notation, suppose that $\delta(G) \geq \beta n$. The key property we will use is the following: if $u, v \in V(G)$ are non-adjacent, then $N(u) \cap N(v)$ is a clique, as otherwise G has an induced C_4. We consider two cases based on the independence number of G.

Case 1: $\alpha(G) \geq \frac{2}{\beta}$. Fix an independent set $I = \{v_1, \dots, v_t\}$ of size $t = \frac{2}{\beta}$. By inclusion-exclusion, we have

$$n \geq |\bigcup_{i \in [t]} N(v_i)| \geq \sum_{i=1}^{t} |N(v_i)| - \sum_{1 \leq i < j \leq t} |N(v_i) \cap N(v_j)| \geq t \cdot \beta n - t^2 \omega(G).$$

Rearranging this gives

$$\omega(G) \geq \frac{\beta n - n/t}{t} = \frac{\beta^2}{4} n,$$

as required.

Case 2: $\alpha(G) \leq \frac{2}{\beta}$. Let $I = \{v_1, \dots, v_t\}$ be a maximum independent set in G. By the maximality of I, every vertex in $V(G) \setminus I$ has a neighbor in I. For $i \in [t]$, let $N^*(v_i)$ be the set of all $u \in V(G) \setminus I$ which is adjacent to v_i but not to any other vertex in I. Then $N^*(v_i)$ is also a clique, because if $u, w \in N^*(v_i)$ are not adjacent then $(I \setminus \{v_i\}) \cup \{u, w\}$ is an independent set larger than I. Moreover, $V(G) \setminus I = \left(\bigcup_{i=1}^{t} N^*(v_i)\right) \cup \left(\bigcup_{1 \leq i < j \leq t} N(v_i) \cap N(v_j)\right)$. Since each of the $t + \binom{t}{2} = \binom{t+1}{2}$ sets in the union on the RHS is a clique, we get that

$$\omega(G) \geq \frac{n - |I|}{\binom{t+1}{2}} \geq \frac{n}{2t^2} \geq \frac{\beta^2}{8} n.$$

■

When proving an induced removal lemma for C_4, we cannot assume that the graph is induced-C_4-free, but only that it has few induced copies of C_4. Hence, we need an approximate version of Theorem 6.1. We can prove such a theorem by combining Theorem 6.1 with the fact that the property of having a clique of size at least ρn is testable. In Section 5 we proved this for independent sets in place of cliques, but the statement for cliques follows by taking complements. More precisely, we will apply Theorem 5.3 to the complement of G. The *density* of a vertex set X is $d(X) := e(X)/\binom{|X|}{2}$.

Lemma 6.2. *For every $\beta, \delta > 0$ there is $\eta = poly(\beta\delta) > 0$ such that the following holds. If an n-vertex graph G has at least $2\beta n^2$ edges and less than ηn^4 induced copies of C_4, then there is $U \subseteq V(G)$ with $|U| \geq \frac{\beta^2}{8} n$ and $d(U) \geq 1 - \delta$.*

Proof sketch. Put $\rho := \frac{\beta^2}{8}$ and $\varepsilon = \frac{\rho\delta^2}{3}$. Note that if G is ε-close to having a clique of size at least ρn, then there is a set $U \subseteq V(G)$ with $|U| \geq \rho n$ and $e(U) \geq \binom{|U|}{2} - \varepsilon n^2 \geq (1-\delta)\binom{|U|}{2}$, so we are done. So suppose (for the sake of contradiction) that G is ε-far from having a clique of size at least ρn. Then by Theorem 5.3 (applied to the complement of G), with probability at least 0.9, a sample $X = \{x_1, \dots, x_q\}$ of $q = \text{poly}(1/\varepsilon) = \text{poly}(\frac{1}{\beta\delta})$ vertices of G satisfies that $\omega(G[X]) < \rho q = \frac{\beta^2}{8}q$. Also, one can show, using Chebyshev's inequality, that with probability at least 0.9, $\frac{e(G[X])}{q^2} \geq (1-\varepsilon)\frac{e(G)}{n^2} \geq \beta$. Hence, with probability at least 0.8, we have both $e(G[X]) \geq \beta q^2$ and $\omega(G[X]) < \frac{\beta^2}{8}q$. By (the contrapositive of) Theorem 6.1, this means that $G[X]$ contains an induced C_4.

Now, let M denote the number of induced-C_4-copies in G. The probability that $X = \{x_1, \dots, x_q\}$ is at least 0.8 (by the above) and at most $M \cdot q^4 \cdot \frac{1}{n^4}$. Hence, $M \geq \frac{0.8n^4}{q^4}$.[27] So taking $\eta := \frac{0.8}{q^4}$, we get a contradiction to the assumption of the lemma. ■

Given Lemma 6.2, our strategy to prove Theorem 3.5 is as follows. Given a graph with few induced copies of C_4, apply Lemma 6.2 repeatedly to find sets $U_1, \dots, U_k$ with $d(U_i) \geq 1-\delta$. By Lemma 6.2, we can continue the process as long as the remaining graph $G - \bigcup_{i=1}^{k} U_i$ has at least $2\beta n^2$ edges. Thus, we obtain a decomposition of $V(G)$ into "almost cliques" and one "almost independent set". We will also need information about the bipartite graph between two almost-cliques.[28]

What can we say about the bipartite graph between two cliques X, Y in an induced-C_4-free graph? The bipartite graph (X, Y) has no induced matching of size 2, because such a matching corresponds to an induced C_4 in G; namely, there are no $x, x' \in X, y, y' \in Y$ such that $y \in N_Y(x) \setminus N_Y(x')$ and $y' \in N_Y(x') \setminus N_Y(x)$. In other words, for every $x, x' \in X$, we have $N_Y(x) \subseteq N_Y(x')$ or $N_Y(x') \subseteq N_Y(x)$. This means that there is an ordering $x_1, \dots, x_m$ of the vertices in X such that $N_Y(x_1) \subseteq N_Y(x_2) \subseteq \dots \subseteq N_Y(x_m)$. Let us call such a bipartite graph a *generalized half-graph.*

As a next step, we need an approximate version of the above structural characterization of a bipartite graph with no bi-induced copies of M_2, the induced matching of size 2. Namely, we want a removal lemma stating that if a bipartite graph has few bi-induced copies of M_2, then it is close to a generalized half-graph.

Lemma 6.3. *For every $\varepsilon > 0$ there is $\eta = poly(\varepsilon) > 0$ such that if a bipartite graph $G = (X, Y)$ has at most $\eta|X|^2|Y|^2$ bi-induced copies of M_2, then G can be made into a generalized half-graph by adding/deleting at most $\varepsilon|X||Y|$ edges.*

One can prove Lemma 6.3 by using (a bipartite version of) Theorem 4.7. In fact, Theorem 4.7 implies a general induced removal lemma in the setting of bipartite graphs, i.e., it works for any forbidden bi-induced subgraph, and not M_2. However, generalized half-graphs are simple enough that one can give a direct argument without using Theorem 4.7. The following proof is due to de Joannis de Verclos.[29]

Proof of Lemma 6.3. Observe that in a generalized half-graph (X, Y), either there is $x \in X$ with

[27]This is a standard double-counting argument used to go from an "abundance statement" in terms of sampling to an "abundance statement" in terms of the number of copies.

[28]Note that we cannot say anything about the bipartite graph between a clique and an independent set, because every split graph is induced C_4-free, so this bipartite graph can be arbitrary.

[29]It is taken from a paper of de Joannis de Verclos showing that the property of being a chordal graph has a polynomial removal lemma.

no neighbors in Y, or there is $y \in Y$ adjacent to all vertices in X.[30]

Now, we run the following process: Initialize $X_0 = X, Y_0 = Y$. For $i \geq 0$, proceed as follows: If there is $x \in X_i$ with at most $\varepsilon|Y|$ neighbors in Y_i, delete all edges between x and Y_i, delete x and continue with the remaining subgraph; i.e., set $X_{i+1} = X_i \setminus \{x\}, Y_{i+1} = Y_i$. Similarly, if there is $y \in Y_i$ which has at most $\varepsilon|X|$ non-neighbors in X_i, add all possible edges between y and X_i, delete y and continue with the remaining subgraph; i.e., set $X_{i+1} = X_i, Y_{i+1} = Y_i \setminus \{y\}$.

Observe that if this process exhausts the entire vertex-set $X \cup Y$, then we turned the graph into a generalized half-graph by adding/deleting at most $2\varepsilon|X||Y|$ edges. So suppose that the process stops at some step i. This means that every $x \in X_i$ has at least $\varepsilon|Y|$ neighbors in Y, and every $y \in Y_i$ has at least $\varepsilon|X|$ neighbors in X. In particular, $|X_i| \geq \varepsilon|X|, |Y_i| \geq \varepsilon|Y|$. Sample a subset $X' \subseteq X_i$ and $Y' \subseteq Y_i$ of size $q := \frac{10}{\varepsilon}\log\frac{1}{\varepsilon}$ each. The probability that some $x \in X'$ has no neighbor in Y' is at most $q(1-\varepsilon)^q \leq \frac{1}{3}$, and similarly, the probability that some $y \in Y'$ has no neighbor in X' is at most $\frac{1}{3}$. It follows that with probability at least $\frac{1}{3}$, (X', Y') is not a generalized half-graph, and hence contains a bi-induced M_2. Using a double counting argument (similar to the one used in the proof of Lemma 6.2), one can now deduce that there are at least $\frac{1}{3q^4}|X_i|^2|Y_i|^2 \geq \frac{\varepsilon^4}{3q^4}|X|^2|Y|^2$ bi-induced copies of M_2. So we can take $\eta = \frac{\varepsilon^4}{3q^4}$. ■

Combining all of the above, we can obtain the following result on the structure of graphs with few induced C_4's.

Lemma 6.4. *For every $\beta, \gamma > 0$ there is $\zeta = poly(\beta, \gamma)$ such that the following holds. If G is an n-vertex graph with at most ζn^4 induced copies of C_4, then there is a partition $V(G) = U_1 \cup \dots \cup U_k \cup W$ with $k \leq poly(1/\beta)$, and there exists a graph G' on $V(G)$, satisfying the following:*

1. *In G', $U_1, \dots, U_k$ are cliques, W is an independent set, and all pairs (U_i, U_j), $1 \leq i < j \leq k$, are generalized half-graphs.*

2. *G' is obtained from G by adding/deleting at most βn^2 edges in total, and at most γn^2 inside $U := U_1 \cup \dots \cup U_k$.*

The distinction in Item 2 – between the total number of edge-changes and their number inside U – is important. The reason is a delicate dependence between the parameters. In the proof of Theorem 3.5, we will start with a given graph G which is ε-far from being induced-C_4-free, and apply Lemma 6.4 to this graph. We will then first clean $G'[U]$, turning it to an induced-C_4-free graph $G''[U]$, by adding/deleting only few edges (βn^2, say). After that, we will argue that one can make few ($\frac{\varepsilon}{2}n^2$, say[31]) changes between U and W to make the entire graph induced-C_4-free; if not, then G'' has many induced-C_4-copies. But to complete the proof, we need to find many induced-C_4-copies in G, not G''. To achieve this, we will need to choose β small enough compared to ε so that the at most $3\beta n^2$ (say) edge-changes made to G to obtain G'' do not ruin all of the induced C_4's we find in G''. Indeed, an added/deleted edge can belong to at most n^2 induced C_4's, so we are fine if we find at least $4\beta n^4$ induced C_4's in G''.

Going back one step – i.e., to the step of cleaning $G'[U]$ to obtain $G''[U]$ – we again have a similar phenomenon. The cleaning scheme guarantees that if there is an induced copy of C_4 in $G''[U]$, then there are many induced-C_4-copies in $G'[U]$. However, we need to find induced copies in G. Consequently, we need to choose γ small enough so that most of the induced-C_4-copies we

[30]Indeed, the first case holds if $N_Y(x_1) = \emptyset$, and the latter if $N_Y(x_1) \neq \emptyset$. Here, $x_1, \dots, x_m$ is the ordering of X with $N_Y(x_1) \subseteq N_Y(x_2) \subseteq \dots \subseteq N_Y(x_m)$.

[31]So that the overall number of changes in all steps is at most εn^2.

find in $G'[U]$ are also present in $G[U]$, as they contain none of the at most γn^2 edges which were added/deleted to turn $G[U]$ into $G'[U]$. This is the reason for having two parameters (i.e., β and γ) in Item 2 of the lemma. In fact, β will be chosen polynomial in ε, while γ will be exponential in ε (i.e., of the form $2^{-\text{poly}(1/\varepsilon)}$), since the cleaning step of handling $G'[U]$ incurs an exponential loss. The bottom line is that at any step of the proof, we need the number of edge-changes in all previous steps to be tiny enough so as to not interfere with the copies that we find in the given step.

Proof sketch of Lemma 6.4. We fix $\delta \ll \gamma$. Now, apply Lemma 6.2 repeatedly, each time finding a set $U \subseteq V(G)$ with $|U| \geq \frac{\beta^2}{8}n$ and $d(U) \geq 1-\delta$, and deleting U from the graph. We can apply Lemma 6.2 as long as the remaining graph has at least $2\beta n^2$ edges. This gives us a partition $V(G) = U_1 \cup \dots \cup U_k \cup W$ such that $|U_i| \geq \frac{\beta^2}{8}n$ and $d(U_i) \geq 1-\delta$ for every $i \in [k]$, and $e(W) \leq 2\beta n^2$. Turn each U_i into a clique and W into an independent set. If there is $1 \leq i < j \leq k$ such that (U_i, U_j) has at least $\eta|U_i|^2|U_j|^2$ bi-induced copies of M_2, then we get many induced-C_4-copies in G. There is a subtlety here (similar to the discussion following the statement of Lemma 6.4): the induced-C_4-copies we find are apriori not in G, but in the graph obtained from G by turning the U_i's into cliques. However, this change consists of adding at most $\delta\binom{|U_i|}{2}$ edges inside U_i, so most of the induced-C_4-copies we find are also present in G, provided that we choose $\delta \ll \eta$. We conclude that for every $1 \leq i < j \leq k$, (U_i, U_j) has less than $\eta|U_i|^2|U_j|^2$ bi-induced copies of M_2. Now use Lemma 6.3 (with parameter γ) to turn each (U_i, U_j) into a generalized half-graph. ∎

In the remainder of this section, we realize the plan outlined in the two paragraphs following Lemma 6.4. First, in the following lemma, we handle the step where we handled the edges between U and W. Recall that at this step, the graph induced by U has already been made induced-C_4-free.

Lemma 6.5. *For every $\varepsilon > 0$ there is $\delta = poly(\varepsilon) > 0$ such that the following holds. Let G be a graph with a vertex-partition $V(G) = U \cup W$ such that $G[U]$ is induced-C_4-free and W is an independent set. If G is ε-far to induced-C_4-free, then G contains at least δn^4 induced copies of C_4.*

Proof. We delete edges between U and W by doing the following for every $w \in W$. Let M_w be a maximal anti-matching in $G[N_U(w)]$; namely, M_w is a maximal collection of non-edges contained in $N_U(w)$. Delete all edges between u and the vertices participating in M. Note that by the maximality of M, in the remaining graph it holds that $N_U(w)$ is a clique. This implies that the remaining graph is induced C_4-free.[32] Since G is assumed to be ε-far from induced C_4-free, we get that

$$\sum_{w \in W} |M_w| \geq \frac{\varepsilon}{2}n^2. \tag{2}$$

Now fix any $w \in W$ and write $M_w = \{x_1y_1, \dots, x_ky_k\}$. Observe that since $G[U]$ has no induced C_4, for every $1 \leq i < j \leq k$ it holds that one of $x_ix_j, x_iy_j, y_ix_j, y_iy_j$ is not an edge. Hence, there are at least $k + \binom{k}{2} \geq k^2/2 = |M_w|^2/2$ induced cherries containing w as the center and two vertices from U.[33] By (2) and Jensen's inequality, the total number of these cherries is at least

$$\frac{1}{2}\sum_{w \in W} |M_w|^2 \geq \frac{1}{2} \cdot |W| \cdot \left(\frac{\varepsilon n^2/2}{|W|}\right)^2 \geq \frac{\varepsilon^2}{8}n^3.$$

[32] Indeed, there is no induced C_4 inside U by assumption, and there is no induced C_4 containing a vertex of W because C_4 has no vertex whose neighborhood is a clique.

[33] Here, by an induced cherry we mean a triple w, x, y with $wx, wy \in E(G)$ and $xy \notin E(G)$. The vertex w is called the center of the cherry.

We can now use the lower bound on the number of induced cherries to count induced copies of C_4. Indeed, by another standard application of Jensen's inequality (this time summing over pairs of vertices in W), we get that there are $\text{poly}(\varepsilon)n^4$ induced copies of C_4. Here we use that W is independent. ■

Proof sketch of Theorem 3.5. Apply Lemma 6.4 to get $U_1, \dots, U_k, W$. We need the following fact:

Claim 6.6. *Consider a generalized half-graph (X, Y). There are partitions $X = X_1 \cup \dots \cup X_t$ and $Y = Y_1 \cup \dots \cup Y_t$ with $t = O(1/\beta)$, such that all pairs (X_i, Y_j) are complete or empty, except for a set of pairs with total weight at most $\beta|X||Y|$.*[34]

The proof is left to the reader. To illustrate the idea, note that if (X, Y) is in fact a half-graph, meaning that $X = \{x_1, \dots, x_m\}, Y = \{y_1, \dots, y_m\}$ and $x_i y_j \in E$ if and only if $i < j$, then partitioning X and Y into equal-size intervals $X_1, \dots, X_t$ and $Y_1, \dots, Y_t$ (where an interval is with respect to the orderings $x_1, \dots, x_m$ and $y_1, \dots, y_m$), we get a partition where (X_i, Y_j) is complete or empty unless $i = j$.

We now take, for every $1 \leq i < j \leq k$, partitions of $\mathcal{P}_{ij}$ and $\mathcal{P}_{ji}$ of U_i and U_j, respectively, as given by the claim. Next, we take the common refinement of all of these partitions. One can show that the resulting partition is β-homogeneous. Note that the size of the partition is roughly $2^k = 2^{\text{poly}(1/\beta)} = 2^{\text{poly}(1/\varepsilon)}$. This is the only place in the argument where we incur an exponential loss. Having found a β-homogeneous partition of $G'[U]$, we can proceed as in the scheme described in Section 3; or, even more similarly, as done in the proof of Theorem 4.9. This allows us to clean the graph $G'[U]$ to make it induced-C_4-free, or else find many induced-C_4-copies in $G'[U]$. As a last step, we apply Lemma 6.5 to handle the edges between U and W. ■

7 Hypergraph regularity and VC-dimension for hypergraphs

In this section we consider the extensions of the notions of regularity and VC-dimension to hypergraphs. For simplicity, we consider 3-uniform hypergraphs, but all material covered in this section extends to higher uniformity.

7.1 Regularity

What is the appropriate notion of regularity for 3-uniform hypergraphs? A natural attempt is as follows: A 3-partite 3-graph $\mathcal{H} = (X, Y, Z)$ is *ε-regular* if for every $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$ with $\frac{|X'|}{|X|}, \frac{|Y'|}{|Y|}, \frac{|Z'|}{|Z|} \geq \varepsilon$, it holds that $|d(X', Y', Z') - d(X, Y, Z)| \leq \varepsilon$, where

$$d(X, Y, Z) := \frac{e(X, Y, Z)}{|X||Y||Z|}$$

is the density of (X, Y, Z). This notion of regularity is called *weak regularity* (for reasons that we shall see shortly). One can indeed prove a regularity lemma with respect to this notion; the statement and its proof are straightforward generalizations of Szemerédi's regularity lemma (Theorem 1.8). The problem, however, is that this notion of regularity is not strong enough to imply a counting lemma. To see this, consider the following key example: Take random bipartite graphs $E \subseteq X \times Y$, $F \subseteq X \times Z$ and $G \subseteq Y \times Z$ (with edge-probability $\frac{1}{2}$, say, though this will not be important). Define the

[34]The weight of (X_i, Y_j) is $|X_i||Y_j|$.

hypergraph $\mathcal{H}$ to consist of all triangles formed by E, F, G, i.e., $xyz \in E(\mathcal{H})$ if $xy \in E, xz \in F, yz \in G$. This hypergraph $\mathcal{H}$ is weakly regular. Indeed, for all linear-sized sets $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$, the density of triangles between (X', Y', Z') is very close to that of (X, Y, Z), because the graphs E, F, G are random. However, $\mathcal{H}$ does not contain a tri-induced copy of every fixed-size 3-partite 3-graph.[35] Indeed, let $\mathcal{K}$ be the 3-partite 3-graph obtained from $K^{(3)}_{2,2,2}$ by deleting one edge. If, by contradiction, $x_1, x_2, y_1, y_2, z_1, z_2$ form a tri-induced copy of $\mathcal{K}$ in $\mathcal{H}$ (with $x_1, x_2 \in X$ and so on), then all pairs $x_i y_j$ $(1 \leq i, j \leq 2)$ belong to E, and similarly all pairs $x_i z_j$ belong to F and all pairs $y_i z_j$ to G. But then this is a copy of $K^{(3)}_{2,2,2}$ instead of an induced copy of $\mathcal{K}$. So we see that the analogue of Lemma 1.7 fails for weak regularity.[36]

Let us now introduce a stronger notion of regularity which does admit a counting lemma. Consider a tripartite graph with parts X, Y, Z consisting of $E \subseteq X \times Y$, $F \subseteq X \times Z$ and $G \subseteq Y \times Z$. Let $\triangle(E, F, G)$ denote the set of triangles in G, i.e., the set of triples $(x, y, z) \in X \times Y \times Z$ with $xy \in E, xz \in F, yz \in G$. Now let $\mathcal{H}$ be a 3-partite 3-graph on X, Y, Z. The *density* of $\mathcal{H}$ with respect to (E, F, G) is defined as

$$d(\mathcal{H} \mid E, F, G) := \frac{|E(\mathcal{H}) \cap \triangle(E, F, G)|}{|\triangle(E, F, G)|}.$$

Namely, the density is the fraction of (E, F, G)-triangles which are edges of $\mathcal{H}$. The definition of regularity in the 3-graph regularity lemma is with respect to this density. That is, we require that for every $E' \subseteq E, F' \subseteq F, G' \subseteq G$, if $\triangle(E', F', G') \geq \varepsilon \triangle(E, F, G)$, it holds that

$$\left| d(\mathcal{H} \mid E', F', G') - d(\mathcal{H} \mid E, F, G) \right| \leq \varepsilon.$$

In fact, the known proof of the 3-graph counting lemma requires a somewhat stronger version of the above: We say that $\mathcal{H}$ is (ε, r)*-regular with respect to* (E, F, G) if for every $E_1, \dots, E_r \subseteq E, F_1, \dots, F_r \subseteq F, G_1, \dots, G_r \in G$ with $\sum_{i=1}^{r} \triangle(E_i, F_i, G_i) \geq \varepsilon \triangle(E, F, G)$, it holds that

$$\left| \frac{|E(\mathcal{H}) \cap \bigcup_{i=1}^{r} \triangle(E_i, F_i, G_i)|}{|\bigcup_{i=1}^{r} \triangle(E_i, F_i, G_i)|} - d(\mathcal{H} \mid E, F, G) \right| \leq \varepsilon.$$

In order to make use of the ε-regularity of $\mathcal{H}$ with respect to (E, F, G), we have to be able to count the (E, F, G)-triangles. To this end, we require that the bipartite graphs E, F, G themselves are regular.[37]

The 3-graph regularity lemma supplies vertex partitions[38] of X, Y, Z as well as pair partitions of $X' \times Y', X' \times Z', Y' \times Z'$ for any choice of vertex-parts $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$. The lemma guarantees that for "most"[39] choices of vertex-parts X', Y', Z' and pair-parts $E \subseteq X' \times Y', F \subseteq$

[35] A tri-induced copy of a 3-partite 3-graph $\mathcal{K} = (A, B, C)$ is defined in the natural way: it is an injection $\varphi : V(\mathcal{K}) \to V(\mathcal{H})$ such that $\varphi(A) \subseteq X, \varphi(B) \subseteq Y, \varphi(C) \subseteq Z$, and for every $(a, b, c) \in A \times B \times C$, $abc \in E(\mathcal{K})$ if and only if $\varphi(a)\varphi(b)\varphi(c) \in E(\mathcal{H})$.

[36] Another classical example to the counting lemma failing is as follows. Take a random tournament T on n vertices, and consider the 3-uniform hypergraph $\mathcal{H}$ on $V(T)$ whose edges are the cyclic triangles in T. It can be shown that $\mathcal{H}$ is weakly-regular (due to the fact that T is random), but any 4 vertices of $\mathcal{H}$ contain at most 2 edges (because any 4 vertices in a tournament contain at most 2 cyclic triangles). In particular, $\mathcal{H}$ does not contain $K^{(3)}_4$ (or even $K^{(3)}_4 - e$).

[37] I.e., they should be regular enough so that we may apply the graph counting lemma (Lemma 1.3). This means that the degree of regularity should be small enough as a function of the densities of E, F, G.

[38] The goal of the vertex partitions is to make the parts of the pair partitions regular.

[39] "Most" means the following: If we sample $(x, y, z) \in X \times Y \times Z$ uniformly at random and consider the unique vertex- and pair-parts containing (x, y, z), then these have the desired property with probability at least $1 - \varepsilon$. In other words, (E, F, G) is weighted by $\frac{|\triangle(E,F,G)|}{|X||Y||Z|}$.

$X' \times Z', G \subseteq Y' \times Z'$, it holds that E, F, G are δ-regular (for a suitable small enough δ) and $\mathcal{H}$ is (ε, r)-regular with respect to E, F, G.[40]

Just as in the graph case, the proof of the 3-graph regularity lemma proceeds via density increment: if $\mathcal{H}$ is not ε-regular with respect to many triples of pair-parts (E, F, G), then one can refine the pair partition and thus increase the energy function. One then needs to apply the graph regularity lemma to the new pair parts to maintain the property that all pair parts are regular. This in turn refines the vertex partition. The repeated applications of graph regularity result in a *Wowzer-type* bound. The wowzer function is the iterated tower function, i.e., $\text{wowzer}(x) = \text{tower}(\text{wowzer}(x-1))$.

7.2 VC-dimension

Recall that a shattered set in a graph is a vertex-set $X = \{x_1, \dots, x_d\}$ such that for every $I \subseteq [d]$, there is a vertex y_I such that $\{i \in [d] : x_i y_I \in E\} = I$. In 3-uniform hypergraphs, a shattered set will consist of pairs instead of vertices; the rest of the definition is very similar. Namely, a set of pairs $\{e_1, \dots, e_d\}$ in a 3-graph $\mathcal{H}$ (so $e_i \in \binom{V(\mathcal{H})}{2}$ for every $1 \leq i \leq d$) is *shattered* if for every $I \subseteq [d]$, there is a vertex $y_I \in V(\mathcal{H})$ such that $e_i \cup \{y_I\} \in E(\mathcal{H})$ if and only if $i \in I$.

As $e_1, \dots, e_d$ are now pairs (instead of vertices), they themselves carry structure, i.e., of a graph. Hence, we can have different definitions of VC-dimension depending on the structure of shattered sets which we are considering.

Definition 7.1.

1. *The* strong VC-dimension *of $\mathcal{H}$ is the maximum size of a shattered set of pairs $e_1, \dots, e_d$ (here there are no restrictions on $e_1, \dots, e_d$).*[41]
2. *The* VC_1-dimension *(also known as* slicewise VC-dimension*) of $\mathcal{H}$ is the maximum size of a shattered set $e_1, \dots, e_d$ which forms a star.*
3. *The* VC_2-dimension *of $\mathcal{H}$ is the maximum size of a shattered set $e_1, \dots, e_d$ which forms a complete bipartite graph.*

For convenience, in what follows we often consider 3-partite 3-graphs (instead of general 3-graphs), but all material applies to general 3-graphs as well.

Fox, Pach and Suk proved that hypergraphs with bounded strong VC-dimension have small homogeneous partitions:

Theorem 7.2 (Fox-Pach-Suk 2019)**.** *If $\mathcal{H} = (X, Y, Z)$ has strong VC-dimension d, then it has an ε-homogeneous equipartition of size at most $(1/\varepsilon)^{O(d)}$. Namely, there are equipartitions $X = X_1 \cup \cdots \cup X_t$, $Y = Y_1 \cup \cdots \cup Y_t$, $Z = Z_1 \cup \cdots \cup Z_t$, where $t \leq (1/\varepsilon)^{O(d)}$, such that for all but at most εt^3 of the triples $(i, j, k) \in [t]^3$ it holds that $d(X_i, Y_j, Z_k) \leq \varepsilon$ or $d(X_i, Y_j, Z_k) \geq 1 - \varepsilon$.*

Thus, strong VC-dimension behaves similarly to the graph case.

Let us now consider VC_1- and VC_2-dimension. Note that if a hypergraph $\mathcal{H}$ has unbounded VC_2-dimension, then it contains a tri-induced copy of every 3-partite 3-graph $\mathcal{K} = (A, B, C)$. Indeed, first map $A \times B$ onto a shattered complete bipartite graph, denoting the mapping by φ, and then, for every $c \in C$, take a vertex $z_c \in V(\mathcal{H})$ which makes an edge precisely with the pairs $\{\varphi(a)\varphi(b) : abc \in E(\mathcal{K})\}$.

[40] For the proof of the counting lemma, the parameter r must also depend on (i.e., be large enough with respect to) the densities of E, F, G.

[41] The term "strong VC-dimension" is not standard. I have not found a better name for this definition.

Thus, in this sense, VC_2-dimension is analogous to the graph case: bounded VC_2-dimension is equivalent to excluding tri-induced copies of a fixed 3-partite 3-graph, just as bounded (graph) VC-dimension is equivalent to excluding bi-induced copies of a fixed bipartite graph. By the same considerations, bounded VC_1-dimension is equivalent to excluding tri-induced copies of a fixed 3-partite 3-graph $\mathcal{K} = (A, B, C)$ where $|A| = 1$. In other words, bounded VC_1-dimension is equivalent to all link graphs having bounded VC-dimension.[42]

It turns out that bounded VC_1-dimension (which is a weaker assumption than bounded strong VC-dimension) also implies the existence of ε-homogeneous vertex-partitions. This was first proved by Chernikov and Towsner, without any quantitative bound on the size of the partition. A double exponential bound $2^{2^{\mathrm{poly}(1/\varepsilon)}}$ was subsequently proved by Terry. Very recently, this was improved to an exponential bound:

Theorem 7.3 (Gishboliner-Shapira-Wigderson)**.** *If $\mathcal{H} = (X, Y, Z)$ has bounded VC_1-dimension, then it has an ε-homogeneous equipartition of size at most $2^{poly(1/\varepsilon)}$.*

A construction by Terry shows that an exponential bound is best possible.[43] The following remains open:

Conjecture 7.4. *If $\mathcal{H} = (X, Y, Z)$ has bounded VC_1-dimension, then there are $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$ with $\frac{|X'|}{|X|}, \frac{|Y'|}{|Y|}, \frac{|Z'|}{|Z|} \geq poly(\varepsilon)$ such that $d(X', Y', Z') \leq \varepsilon$ or $d(X', Y', Z') \geq 1 - \varepsilon$.*

Proof sketch of Theorem 7.3. For simplicity, suppose that $|X| = |Y| = |Z| = n$. Fix any pair $(x, y) \in X \times Y$. Consider the link $L_{\mathcal{H}}(x)$, which is a bipartite graph between Y and Z. As $\mathcal{H}$ has bounded VC_1-dimension, $L_{\mathcal{H}}(x)$ has bounded VC-dimension. Hence, by (the bipartite version of) Lemma 4.4, there is a partition $Y = Y_1^{(x)} \cup \cdots \cup Y_s^{(x)}$, $s = \mathrm{poly}(1/\varepsilon)$, such that two vertices y_1, y_2 in the same part satisfy $|N_Z(y_1) \triangle N_Z(y_2)| \leq \varepsilon n$, where the neighborhoods are in $L_{\mathcal{H}}(x)$. In other words, $|N_Z(x, y_1) \triangle N_Z(x, y_2)| \leq \varepsilon n$, where the neighborhoods are in $\mathcal{H}$. For simplicity, suppose that the partition $Y_1^{(x)} \cup \cdots \cup Y_s^{(x)}$ is an equipartition (this can be easily arranged by allowing one exceptional part). Also, without loss of generality, suppose that $y \in Y_1^{(x)}$. Pick any $y' \in Y_1^{(x)}$, and now consider $L_{\mathcal{H}}(y)$, which is a bipartite graph between X and Z. By the same argument as above, we get an equipartition $X = X_1^{(y')} \cup X_s^{(y')}$ such that two vertices x_1, x_2 in the same part satisfy $|N_Z(x_1, y') \triangle N_Z(x_2, y')| \leq \varepsilon n$. Without loss of generality, $x \in X_1^{(y')}$. Now, for every $x' \in X_1^{(y')}$, we have, by the triangle inequality:

$$|N_Z(x', y') \triangle N_Z(x, y)| \leq |N_Z(x', y') \triangle N_Z(x, y')| + |N_Z(x, y') \triangle N_Z(x, y)| \leq 2\varepsilon n.$$

Also, the number of choices for (x', y') is $(n/s)^2$. Summarizing, for every pair[44] $(x, y) \in X \times Y$, there are at least $(n/s)^2$ pairs $(x', y') \in X \times Y$ with $|N_Z(x', y') \triangle N_Z(x, y)| \leq 2\varepsilon n$.

Now sample $f_1, \ldots, f_r \in X \times Y$ uniformly at random, where $r = s^2 \log \frac{1}{\varepsilon}$, and define $E_i := \{(x, y) \in X \times Y : |N_Z(x, y) \triangle N_Z(f_i)| \leq 2\varepsilon n\}$. For every $i \in [r]$, very two pairs in F_i have the same

[42]The link $L_{\mathcal{H}}(x)$ of a vertex x is the graph $\{yz : xyz \in E(\mathcal{H}\}$.

[43]This construction is as follows: Partition X, Y into equal-sized parts $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$, respectively, where $k = (1/\varepsilon)^{0.1}$, say. For each $i = 1, \ldots, k$, take a uniformly random subset $Z_i \subseteq Z$ and add all edges in $X_i \times Y_i \times Z_i$ (the sets $Z_1, \ldots, Z_k$ are chosen independently). One can show that any ε-homogeneous partition of this hypergraph has size at least $2^{(1/\varepsilon)^{\Omega(1)}}$.

[44]In fact, we can only guarantee this for almost every pair, because of the aforementioned exceptional sets. But we ignore this technicality.

neighborhood in Z, up to an error of $4\varepsilon n$ (by the triangle inequality). Also, for a given $(x, y) \in X \times Y$, we have

$$\mathbb{P}[(x, y) \notin E_1 \cup \cdots \cup E_r] \leq \left(1 - \frac{1}{s^2}\right)^r \leq \varepsilon.$$

In conclusion, we get a partition $X \times Y = E_0 \cup E_1 \cup \cdots \cup E_r$ such that $|E_0| \leq \varepsilon n^2$, and for every $i \in [r]$ and $(x, y) \in E_i$, $|N_Z(x, y) \triangle N_Z(f_i)| \leq 2\varepsilon n$.

Now, let $Z_i := N_Z(f_i)$, and let $\mathcal{P}_Z$ be the Venn diagram of the sets $Z_1, \ldots, Z_r$. This is a partition of Z into at most $2^r = 2^{\text{poly}(1/\varepsilon)}$ sets.[45] We expect a homogeneous behavior between each E_i ($1 \leq i \leq r$) and a typical part of $\mathcal{P}_Z$.

The proof proceeds by repeating the above argument for $X \times Z$ and $Y \times Z$. In the former case we obtain a partition of $X \times Z$ and a partition $\mathcal{P}_Y$ of Y, and in the latter case we obtain a partition of $Y \times Z$ and a partition $\mathcal{P}_X$ of X. One can now show that $(\mathcal{P}_X, \mathcal{P}_Y, \mathcal{P}_Z)$ is an ε-homogeneous partition of $\mathcal{H}$.[46] We omit the details. ■

Moving to VC$_2$-dimension, what can we say about a hypergraph with bounded VC$_2$-dimension? Recall the construction described in Section 7.1 (arising from random graphs). This construction has bounded VC$_2$-dimension but no ε-homogeneous partition (even for $\varepsilon = 0.49$) of size independent of n. So bounded VC$_2$-dimension does not imply the existence of (bounded-size) ε-homogeneous vertex partitions. Observe, however, that this construction does have a homogeneous *pair partition*: Partition $X \times Y$ into $E_1 := E, E_2 := X \times Y \setminus E$, and similarly partition $X \times Z$ into F_1, F_2 and $Y \times Z$ into G_1, G_2. Then for every $i, j, k = 1, 2$, the hypergraph $\mathcal{H}$ is homogeneous (i.e., complete or empty) over $\triangle(E_i, F_j, G_k)$. It turns out that this is a general phenomenon:

Theorem 7.5 (Chernikov-Towsner 2020)**.** *If $\mathcal{H} = (X, Y, Z)$ has bounded VC$_2$-dimension, then there are equipartitions $X \times Y = E_1 \cup \cdots \cup E_t$, $X \times Z = F_1 \cup \cdots \cup F_t$, $Y \times Z = G_1 \cup \cdots \cup G_t$, where t depends only on ε, such that for all but εt^3 of the triples (E_i, F_j, G_k) it holds that $d(\mathcal{H} \mid E_i, F_j, G_k) \leq \varepsilon$ or $d(\mathcal{H} \mid E_i, F_j, G_k) \geq 1 - \varepsilon$.*

Theorem 7.5 can be deduced from the hypergraph regularity lemma, as follows: Taking a regular partition, one can show that for a regular triple (E_i, F_j, G_k), the density of $\mathcal{H}$ over $\triangle(E_i, F_j, G_k)$ is at most ε or at least $1 - \varepsilon$. Indeed, otherwise $\mathcal{H}$ contains a tri-induced copy of every fixed 3-partite 3-graph, by a counting lemma analogous to Lemma 1.7.

In fact, the result of Chernikov and Towsner is more general. For each uniformity $k \geq 2$ and $1 \leq \ell \leq k - 1$, they define a suitable notion of VC$_\ell$-dimension, and show that a k-graph with bounded VC$_\ell$-dimension has an *ε-homogeneous partition of uniformity ℓ*, i.e., a partition of all ℓ-sets.

Problem 7.6. *Does Theorem 7.5 hold with $t = poly(1/\varepsilon)$?*

[45] This step is where the exponential bound in Theorem 7.3 comes from.

[46] More precisely, ε'-homogeneous for some $\varepsilon' = \varepsilon^c$, $c > 0$ constant.